

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/00, 15/52, 15/70, 1/21, C07K 14/195</b>		<b>A2</b>	(11) International Publication Number: <b>WO 96/14405</b>
			(43) International Publication Date: 17 May 1996 (17.05.96)
(21) International Application Number: <b>PCT/US95/15327</b>		(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).	
(22) International Filing Date: 3 November 1995 (03.11.95)			
(30) Priority Data: 08/334,645 4 November 1994 (04.11.94) US			
(71) Applicant: MOLECULAR BIOLOGY RESOURCES, INC. [US/US]; 5520 W. Burleigh Street, Milwaukee, WI 53210 (US).			
(72) Inventors: MUELLER, Reinhold, D.; 1930 North 86th Street, Wauwatosa, WI 53226 (US). SKOWRON, Piotr, M.; 3533 Salerno Court #8, Middleton, WI 53562 (US). SWAMINATHAN, Neela; 3026 Bosshard Drive, Madison, WI 53711 (US). PIEHL, Richard, F.; 2317 East Menlo Boulevard, Shorewood, WI 53211 (US).			
(74) Agents: POCHOPIEN, Donald, J. et al.; 6300 Sears Tower, 233 South Wacker Drive, Chicago, IL 60606-6402 (US).			

**Published***Without international search report and to be republished upon receipt of that report.**With an indication in relation to a deposited microorganism furnished under Rule 13<sup>bis</sup> separately from the description.**Date of receipt by the International Bureau:  
29 December 1995 (29.12.95)***(54) Title: BIOLOGICALLY ACTIVE FRAGMENTS OF THERMUS FLAVUS DNA POLYMERASE****(57) Abstract**

The present invention is directed to a DNA encoding a biologically active fragment of a thermostable, full length DNA polymerase I enzyme of *Thermus flavus*. More particularly, the invention is directed to a DNA encoding an approximately 63,000 dalton DNA polymerase that lacks 274 amino acids from the N-terminus of the approximately 94,000 dalton *T. flavus* DNA polymerase I, and to the protein encoded thereby which has been designated the *T. flavus* DNA polymerase I exo fragment. The enzyme fragments are useful in DNA sequencing, Thermal Cycle Labeling, Polymerase Chain Reaction, and other molecular biological applications.

# *FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## BIOLOGICALLY ACTIVE FRAGMENTS OF *THERMUS FLAVUS* DNA POLYMERASE

### BACKGROUND OF THE INVENTION

#### A. Field of the Invention

- 5 The present invention relates to an isolated and purified DNA that encodes a thermostable DNA polymerase. Additionally, the present invention relates to a recombinant and thermostable DNA polymerase and to fragments thereof, all having enhanced polymerase activity, and to methods for producing the DNA polymerase and fragments. The present invention
- 10 further relates to recombinant fragments having decreased exonuclease activity. The thermostable recombinant polymerases of the present invention are useful because they are capable of providing enhanced polymerase activity in bio-applications, such as in the polymerase chain reaction (PCR), in DNA amplification and in thermal cycle labeling (TCL).

#### 15 B. Background

- The burgeoning field of biotechnology was revolutionized by recombinant DNA technology, and DNA polymerase enzymes are an indispensable tool used in many modern molecular *in vitro* recombinant DNA biological applications, such as in DNA sequencing; DNA cycle sequencing;
- 20 Polymerase Chain Reaction (PCR) and its many variations (see, e.g., Erlich *et al.*, Current Communications in Molecular Biology: Polymerase Chain Reaction. Cold Spring Harbor Press, Cold Spring Harbor (1989); Innis *et al.*, PCR protocols: A guide to methods and applications. Academic Press, San Diego (1990)); Thermal Cycle Labeling (TCL) (Mead and Swaminathan, U.S.
- 25 Patent App. Ser. No. 08/217,459, filed March 24, 1994; PCT App. No. US94/03246, filed March 24, 1994); Random Primer Labeling (RPL); Ligase Chain Reaction (LCR) (Wiedmann *et al.*, *PCR Methods and Applications 3*: S51-S64 (1994)); and other applications.

- 2 -

To date scientists have reported more than 40 different DNA polymerases, and have reported DNA sequence information for some DNA polymerase genes. Amino acid sequence information has been deduced from the reported genes, and comparison of amino acid sequences has resulted in the placement of reported polymerase genes into four major families: namely, A, B, C, and X. Family A contains *E. coli* DNA polymerase I, an enzyme involved in repair of DNA and in replication during fast growth. Family B includes *E. coli* DNA polymerase II. Family C polymerases include *E. coli* DNA polymerase III, the major replication enzyme. The fourth group, Family X, contains enzymes such as the eukaryotic DNA polymerase  $\beta$  and eukaryotic terminal transferases (Ito and Braithwaite, *Nucleic Acids Res.* 19: 4045-4057 (1991)). The breakdown of DNA polymerases into families has been helpful for the understanding of fundamental biological processes and for the selection of enzymes for particular molecular biological applications.

DNA polymerase I (pol I) (Family A) enzymes have proved to be very useful for DNA sequencing applications, PCR, TCL, and other applications known in the art. Structure-function relationship studies indicate that known DNA pol I molecules share a similar modular organization. A 5'  $\rightarrow$  3' exonuclease function is located in the N-terminal one-third of the enzyme. The remainder of the molecule forms one domain which is further classified into functional sub-domains. Adjacent to the 5'  $\rightarrow$  3' exonuclease domain lies a 3'  $\rightarrow$  5' exonuclease sub-domain, followed by a polymerase sub-domain (Blanco et al., *Gene* 100:27-38 (1991)).

In addition to classifying DNA polymerase enzymes into the above families, it is also useful to classify such polymerases as mesophilic (purified from mesophilic organisms) or thermophilic (purified from thermophilic organisms) in origin. DNA polymerases of mesophilic organisms were discovered earlier and have been more extensively studied than their thermophilic counterparts. As early as the 1950's, isolation and purification protocols for DNA polymerase I from mesophilic bacteria (e.g.,



*E. coli*) and some of their phages were developed and have since been modified. See, e.g., Bessman *et al.*, *J. Biol. Chem.* 233:171-177 (1958); Buttin and Kornberg, *J. Biol. Chem.* 241:5419-5427 (1966). The DNA polymerases studied most extensively are the DNA polymerase I enzymes isolated from *E. coli* and the bacteriophage T7 DNA polymerase.

The DNA polymerases of mesophilic origin are useful in many biological applications, such as in certain DNA sequencing applications. However, many important applications (e.g., polymerase chain reaction (PCR) applications and thermal cycle labeling (TCL)) require thermal cycling to repeatedly denature template DNA and/or RNA and their extension products. Because mesophilic DNA polymerases do not withstand the high temperatures or thermal cycling of these applications, thermostable DNA polymerases enjoy significant advantages over mesophilic DNA polymerases in such applications.

The discovery and study of such thermostable DNA polymerases -- from thermophilic bacteria -- has been a much more recent phenomenon. See, e.g., Uemori *et al.*, *J. Biochem.* 113: 401-410 (1993); Uemori *et al.*, *Nucleic Acids Res.* 21: 259-265 (1993); Lawyer *et al.* *J. Biol. Chem.* 264: 6427-6437 (1989); Kaledin *et al.*, *Biokhimiya* 45:644-651 (1980); Chien *et al.*, *J. Bacteriol.* 127:1550-1557 (1976); Gelfand *et al.*, U.S. Patent Nos. 4,889,818 and 5,079,352; Burke *et al.*, U.S. Patent No. 5,108,892.

Perhaps the best-studied thermostable DNA polymerase, derived from *Thermus aquaticus*, is called Taq pol I. A number of routes have been taken in attempts to clone the Taq DNA pol I gene. (See, e.g., Lawyer *et al.* (1989); Gelfand *et al.*, U.S. Patent No. 5,079,352 (1992) (purification to approx. 200,000 units/mg reported); Lawyer *et al.*, *PCR Methods and Applications* 2:275-287 (1993) (purification to 292,000 units/mg reported); Engelke *et al.*, *Anal. Biochem.* 191:396-400 (1990); Sagner *et al.*, *Gene* 97:119-123 (1991)).

As explained above, in addition to possessing useful DNA polymerase activity, a number of DNA polymerase I holoenzymes possess exonuclease activities, which for many biological applications are undesirable. Therefore, modified DNA polymerase enzymes having reduced exonuclease activities are desirable. Through deletion of the 5' one-third of DNA polymerase I genes, or the proteolytic cleavage and subsequent removal of the portion of the holoenzyme encoded thereby, scientists have created DNA pol I fragments retaining polymerizing activity, but having reduced 5' → 3' exonuclease activity. (See, e.g., Joyce and Grindley, *Proc. Natl. Acad. Sci.* 80:1830-1834 (1983) (the Klenow-Fragment of the *E. coli* DNA polymerase enzyme); Lawyer *et al.*, *J. Biol. Chem.* 264:6427-6437 (1989), Gelfand *et al.*, U.S. Patent No. 5,079,352 (1992), Lawyer *et al.*, *PCR Methods and Applications* 2:275-287 (1993) (the Stoffel fragment of the *T. aquaticus* (Taq) DNA polymerase enzyme, reportedly purified to a specific activity of 369,000 units/mg); and Barnes, *Gene* 112:29-35 (1992) (the KlenTaq DNA polymerase).)

In addition to Taq DNA polymerases, other thermophilic DNA polymerases reportedly have been cloned and expressed in *E. coli*. Uemori *et al.* reportedly expressed DNA polymerases from *Bacillus caldoteanax* (*J. Biochem.* 113:401-410 (1993)) and *Pyrococcus furiosus* (*Nucleic Acids Res.* 21:259-265 (1993)).

DNA polymerases from other bacteria of the genus *Thermus* have been reported. A method of recovering a thermostable DNA polymerase from cultured *Thermus thermophilus* is reported in U.S. Patent No. 5,242,818 to Oshima *et al.* (1993). The purported purification of native *Thermus flavus* DNA polymerase with an apparent molecular weight of 66,000 daltons was described by Kaledin *et al.*, *Biokhimiya* 46:1576-1584 (1981)). In one application, Kainz *et al.*, *Anal. Biochem.* 202:46-49 (1992), reported the amplification of a 10.9 kb fragment and a 15.6 kb fragment from phage

lambda DNA with Hot Tub (*T. flavus*) polymerase (Amersham, Arlington Heights, IL), but not with Taq polymerases. The rapid filter assay of Sagner et al., *Gene* 97:119-123 (1991) has allowed Akhmetzjanov and Vakhitov to identify a purported *T. flavus* (strain and origin unidentified) DNA polymerase I gene and to determine the DNA sequence of this gene (*Nucleic Acids Res.* 20:5839 (1992)). There is no report of the expression of an active DNA polymerase encoded by the purported *Thermus flavus* DNA polymerase I (Tfl DNA pol I) gene characterized by Akhmetzjanov and Vakhitov. Native *T. flavus* (Tfl) DNA polymerase I is commercially available, e.g., from Molecular Biology Resources, Inc. (Milwaukee, WI, Catalogue #1112-01).

The different reports of thermostable DNA polymerases and their derivatives suggest these enzymes possess different, unpredictable properties that may be advantageous or detrimental, depending on the biological application in which the DNA polymerase is to be employed. For example, *Thermus thermophilus* DNA polymerase I was reported to have a significant reverse transcriptase activity. In the same reaction tube, in successive steps, the reverse transcriptase function allows the production of double stranded DNA from RNA and then the DNA polymerase function is used to amplify this cDNA. Myers and Gelfand, *Biochemistry* 30:7661-7666 (1991).

The KlenTaq DNA polymerase is an example of an enzyme fragment with important properties differing from the Taq holoenzyme. The KlenTaq DNA polymerase reportedly has a roughly two-fold lower PCR-induced relative mutation rate than Taq polymerase holoenzyme. However, more units of KlenTaq are needed to obtain PCR products similar to those generated with Taq DNA pol I.

Similarly, Lawyer et al. (1993) reported that *T. aquaticus* DNA polymerase I fragments possessed greater thermostability and were active over a broader  $Mg^{++}$  -range than the corresponding holoenzyme. Because of its

- 6 -

broader range of magnesium ion concentration, the Stoffel fragment has been used in multiplex PCR, where more than two primers must anneal to the template. The thermostability of the Stoffel fragment makes this enzyme a better choice when GC-rich templates are amplified. It is desirable to purify and isolate additional DNA polymerase enzymes and derivatives, to take advantage of the unique but unpredictable properties that such molecules may have.

There remains a need in the art for new, thermostable DNA polymerase enzymes for use in the expanding universe of molecular biological applications. More particularly, there exists a need for thermostable DNA polymerase enzymes having high purity, high DNA polymerase specific activity, low levels of exonuclease activity, and possessing high fidelity (low mutation frequencies) and high processivity when used in DNA amplification protocols.

An object of the present invention is to provide polymerase enzyme preparations of greater purity, quantity, DNA polymerase specific activity, and processivity than has heretofore been possible. A further object is to eliminate the need and expense of culturing of large volumes of thermophilic bacteria at high temperatures that is associated with preparing thermostable polymerase enzyme preparations. Yet another object is to provide a recombinant polymerase possessing reduced exonuclease activities, as compared to the currently available native holoenzyme.

### SUMMARY OF THE INVENTION

The present invention relates to the cloning and expression of a gene encoding a thermostable DNA polymerase, the purification of a recombinant thermostable DNA polymerase encoded by the gene, and applications for using the polymerase. The gene of the *Thermus flavus* DNA polymerase I (Tfl DNA pol I), was cloned and expressed in *Escherichia coli*. The purified recombinant *T. flavus* DNA polymerase enzyme is shown to be

- 7 -

thermostable and have a molecular weight of about 90,000 to 100,000 daltons. The DNA sequence of the Tfl DNA pol I gene, including flanking sequences, was determined and the coding sequence of the recombinant enzyme was mapped within this gene. A Tfl DNA Pol I gene fragment also was expressed in *E. coli*, the purified recombinant protein products ("exo<sup>-</sup> fragment") lacking 274 and 275 amino acids from the N-terminus of the Tfl DNA pol I holoenzyme. This Tfl exo<sup>-</sup> fragment has very low 3' → 5' and 5' → 3' exonuclease activities. Numerous properties of and applications for the recombinant enzymes are described.

In one aspect, this invention provides purified polynucleotides (e.g. DNA sequences and RNA transcripts thereof) encoding a thermostable polypeptide having DNA polymerase activity. Preferred DNAs include the *Thermus flavus* DNA pol I gene comprising nucleotides 301 to 2802 of SEQ ID NO: 1; the *Thermus flavus* DNA pol I exo<sup>-</sup> fragment gene comprising nucleotides 1 to 1791 of SEQ ID NO: 3; the DNA comprising nucleotides 112-1791 of SEQ ID NO: 3; a portion of the insert of plasmid pTFLRT4 (ATCC Accession No. 69633), said portion encoding a thermostable polypeptide having DNA polymerase activity; a portion of the insert of plasmid p21EHcM1.1, (ATCC Accession No. 69632), said portion encoding a thermostable polypeptide having DNA polymerase activity; fragments or portions of these DNAs that encode thermostable polypeptides having DNA polymerase activity; and variants of these DNAs that encode thermostable polypeptides having DNA polymerase activity.

In another aspect, this invention provides DNA sequences such as those described above operatively linked to a promoter sequence, a cloning vector, an expression vector, or combinations thereof.

In related aspects, the invention provides novel plasmids and vectors. For example, the invention provides a plasmid pTFLRT4, having ATCC Accession No. 69633; and a plasmid p21EHcM1.1, having ATCC Accession No. 69632. The invention also provides a vector that includes

nucleotides 301 to 2802 of SEQ ID NO:1, the nucleotides encoding a polypeptide having thermostable DNA polymerase activity; and a vector that includes nucleotides 112 to 1791 of SEQ ID No:3, the nucleotides encoding a polypeptide having thermostable DNA polymerase activity.

- 5 In related aspects, the invention provides a vector having at least one insert consisting essentially of nucleotides 301 to 2802 of SEQ ID NO: 1, the nucleotides encoding a polypeptide having thermostable DNA polymerase activity. The invention further provides a vector having at least one insert consisting essentially of nucleotides 112 to 1791 of SEQ ID NO:  
10 3, the nucleotides encoding a polypeptide having thermostable DNA polymerase activity.

- The present invention is also directed to host cells, such as prokaryotic and eukaryotic cells, that have been stably transformed with DNAs vectors, or plasmids of the invention. Another aspect of the invention  
15 is directed to such transformed host cells that are capable of expressing a thermostable polypeptide encoded by the DNAs, the peptide having DNA polymerase activity.

- In another aspect, this invention provides purified thermostable polypeptides having DNA polymerase activity. Preferred peptides include a  
20 *Thermus flavus* DNA polymerase I holoenzyme substantially free of other *Thermus flavus* proteins; a polypeptide having the amino acid sequence of SEQ ID NO: 2; a fragment of a *Thermus flavus* DNA polymerase I holoenzyme, including a fragment with reduced exonuclease activity as compared to the holoenzyme, and also including a fragment having the amino  
25 acid residues 1-560 or 2-560 of the amino acid sequence shown in SEQ ID NO: 5; a fragment encoded by the insert of plasmid p21EHcM1.1, having ATCC Accession No. 69632; fragments of the above peptides that retain DNA polymerase activity; and variants of the above peptides that retain DNA polymerase activity.

In another aspect, this invention provides methods for purifying a thermostable polypeptide having DNA polymerase activity including the steps of transforming a host cell with a DNA of the present invention to create a transformed host cell; cultivating the transformed host cell under conditions that promote expression of a thermostable polypeptide encoded by the DNA, the polypeptide having DNA polymerase activity; and purifying the thermostable polypeptide with a monoclonal antibody that is cross-reactive with the thermostable polypeptide. In one preferred method, the cross-reactive monoclonal antibody has specificity for a *Thermus aquaticus* DNA polymerase and/or for a *Thermus flavus* DNA polymerase.

In another preferred method, commercially available chromatography columns are used to purify the expressed polypeptide.

In another aspect, this invention provides methods of purifying a thermostable polypeptide having DNA polymerase activity. One such method includes the steps of expressing the thermostable polypeptide in a host cell, the polypeptide having an amino acid sequence encoded by a DNA of the present invention; lysing the cell to create a suspension containing the thermostable polypeptide, as well as host cell proteins and cell debris; contacting a soluble portion of the suspension with an antibody that is immunologically cross-reactive with the thermostable polypeptide under conditions wherein the antibody binds to the thermostable polypeptide to form an antibody-polypeptide complex; isolating the antibody-polypeptide complex; and separating the thermostable polypeptide from the isolated antibody-polypeptide complex to provide a purified thermostable polypeptide. Preferably, such a method further includes the steps of heating the suspension to denature the host cell proteins; and centrifuging the suspension to remove the cell debris and denatured host cell proteins. In more preferred methods, the immunologically cross-reactive antibody is a monoclonal antibody, such as a monoclonal antibody that is immunologically cross-reactive with *Thermus aquaticus* DNA polymerase I and/or *Thermus flavus* DNA polymerase I. This

- 10 -



preferred method is exemplified herein using the monoclonal antibody purified from a hybridoma designated hybridoma 7B12.

In another aspect, this invention provides methods of using the DNA constructs of the invention to produce recombinant thermostable polypeptides having DNA polymerase activity. One such method involves using a DNA encoding a DNA polymerase enzyme to generate active fragments of the DNA polymerase enzyme, including the steps of: deleting a portion of the DNA to create a modified DNA; expressing the modified DNA to produce a DNA polymerase enzyme fragment; purifying the DNA polymerase enzyme fragment; assaying the DNA polymerase enzyme fragment for DNA polymerase activity; and selecting a DNA polymerase enzyme fragment having DNA polymerase activity; wherein the DNA is selected from among the DNAs described herein.

In another aspect, this invention provides methods for using the proteins of the invention in biological applications, such as DNA sequencing; amplification of DNA and/or RNA sequences; polymerase chain reaction (PCR); thermal cycle labeling (TCL); universal thermal cycle labeling (UTCL); ligase chain reaction (LCR); and other applications or processes that would be apparent to those skilled in the art.

In yet another aspect, this invention provides kits for using the proteins of the invention in various biological applications, such as kits for labeling DNA.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

FIGURES 1A and 1B graphically depict the cloning strategy: (1A) for the gene encoding the Tfl DNA pol I holoenzyme; and (1B) for the DNA encoding the exo<sup>-</sup> fragment of *T. flavus* DNA polymerase I. The abbreviations used are: B: BamHI, RI: EcoRI, RV: EcoRV, Hc: HincII, P lacZ: promoter of the lacZ gene, S: Sall, and X: XbaI. Jagged lines (  ) represent vector DNA; straight horizontal (  ) lines represent Tfl



- 11 -

insert DNA; dark and light shaded rectangles depict Tfl DNA pol I gene sequences. The graphical depictions are not drawn to scale, and not all available restriction sites are shown in all steps.

FIGURE 2 depicts the DNA sequence and the deduced amino acid sequence for the Tfl DNA pol I holoenzyme coding sequence and for 5' untranslated and 3' untranslated sequences. The circled amino acid (Glu<sub>239</sub>) is the first amino acid believed to be translated during translation of plasmid p21EHcM1.1, encoding the Tfl exo' fragment. The boxed amino acid (Leu<sub>275</sub>) is the amino acid determined to be the first amino acid of the purified and isolated major Tfl exo' fragment. An asterisk (\*) indicates the stop codon TAG.

FIGURE 3 is a comparison of deduced amino acid sequences from the *Thermus flavus* DNA polymerase I of this invention (MBR TFL); *Thermus aquaticus* DNA polymerase I (TAQ) reported in Lawyer *et al.*, *J. Biol. Chem.* 264:6427-6437 (1989)); and purported *Thermus flavus* DNA polymerase I (A&V TFL) described in Akhmetzjanov and Vakhitov, *Nucleic Acids Res.* 20:5839 (1992). The sequences were aligned to maximize homology. Conservative differences between the amino acid sequences are indicated with asterisks (\*) and non-conservative differences are indicated with arrowheads (^).

FIGURE 4 depicts double-stranded DNA sequence of the *T. flavus* DNA pol I gene, including 5' untranslated and 3' untranslated sequences. Lower case letters indicate untranslated sequences, upper case letters represent the coding sequence. The start codon (ATG) of Tfl DNA pol I is at positions 301-303, and the stop codon is at positions 2803-2805. The positions in the sequence that correspond to synthetic primers used in sequencing reactions have been indicated with boxes. The sequence of the 2-4 fragment is underlined with an arrow (<—————>).

FIGURE 5 depicts the relative DNA polymerase enzymatic activity, at different buffered pH levels, of native *Thermus flavus* holoenzyme (nTfl Holo: empty squares); recombinant *Thermus flavus* holoenzyme (rTfl Holo: diamonds); *Thermus flavus* exo' fragment (Tfl exo': circles); *T. aquaticus* DNA pol I (AmpliTaq: crossed boxes); and the Taq enzyme Stoffel fragment (Stoffel: triangles).

FIGURE 6A depicts the relative DNA polymerase enzymatic activity, at different concentrations of  $MgCl_2$ , of native *Thermus flavus* holoenzyme (nTfl Holo: empty squares); recombinant *Thermus flavus* holoenzyme (rTfl Holo: diamonds); *Thermus flavus* exo' fragment (Tfl exo': circles); and *T. aquaticus* DNA pol I Stoffel fragment (Stoffel: triangles).

FIGURE 6B depicts the relative DNA polymerase enzymatic activity, at different concentrations of  $MnCl_2$ , of native *Thermus flavus* holoenzyme (nTfl Holo: empty squares); recombinant *Thermus flavus* holoenzyme (rTfl Holo: diamonds); *Thermus flavus* exo' fragment (Tfl exo': circles); and *T. aquaticus* DNA pol I Stoffel fragment (Stoffel: triangles).

FIGURE 7A depicts the relative DNA polymerase enzymatic activity, at different temperatures, of native *Thermus flavus* holoenzyme (nTfl Holo: open boxes); recombinant *Thermus flavus* holoenzyme (rTfl Holo: diamonds); *Thermus flavus* exo' fragment (Tfl exo': circles); and *T. aquaticus* DNA pol I Stoffel fragment (Stoffel: triangles).

FIGURE 7B photographically depicts the relative quantities of PCR amplification product generated after 25, 30, and 35 reaction cycles, using 10 units of Tfl exo' fragment (E) or Stoffel fragment (S) as the PCR DNA polymerase. The far right lane depicts the PCR amplification product generated after 35 reaction cycles using 1.1 unit of Tfl exo' fragment.

FIGURE 8 depicts enzymatic stability in thermal cycling (relative DNA polymerase enzymatic activity after different numbers of PCR cycles), of native *Thermus flavus* holoenzyme (nTfl Holo: empty squares);

- 13 -

*Thermus flavus* *exo'* fragment (Tfl *exo'*: circles); *T. aquaticus* DNA pol I (AmpliTaq: crossed squares); and *T. aquaticus* DNA pol I Stoffel fragment (Stoffel: triangles).

FIGURE 9 photographically depicts the purity of purified *E. coli* DNA polymerase I (Eco Pol I, control), recombinant *Thermus flavus* holoenzyme (rTfl Holo), and *Thermus flavus* *exo'* fragment (Tfl *exo'*) on a 12.5% SDS-PAGE gel stained with silver.

FIGURES 10A, 10B, 10C, and 10D photographically depict portions autoradiographs of sequencing gels showing DNA sequence obtained with the indicated polymerases substituted into the SEQUAL™ or the Cycle SEQUAL™ DNA Sequencing Kit. Abbreviations: recombinant *Thermus flavus* holoenzyme (Tfl Holo); *Thermus flavus* *exo'* fragment (Tfl *exo'*); *T. aquaticus* DNA pol I holoenzyme (AmpliTaq); and the Taq enzyme Stoffel fragment (Taq Stoffel).

#### DETAILED DESCRIPTION OF THE INVENTION

This application describes the isolation and characterization of the gene coding for *Thermus flavus* (ATCC Accession No. 33923) DNA polymerase I (Tfl DNA pol I) and having homology to the family A enzymes described above. Also described is the expression of this gene in *E. coli* and the purification and characterization of the recombinant DNA polymerase. The cloning and expression of an active fragment of the *Thermus flavus* DNA polymerase gene is also described, and the gene fragment and expressed peptides are characterized. Recombinant vectors and host cells are also described. Additionally, methods and kits are described that involve the DNAs and proteins of the present invention. Thus, as the discussion below details, the present invention has several aspects.

As a first step in the generation of the DNAs and polypeptides of the present invention, native *T. flavus* DNA polymerase I was purified and

- 14 -

isolated from *T. flavus* cells (ATCC Accession No. 33923) and digested with trypsin, and amino acid sequence information was obtained from one of the reaction products (i.e. from a trypsin digest protein fragment). (See Example 1.) Additionally, a *Thermus flavus* genomic library was constructed in phage  
5  $\lambda$  Dash II and amplified. (See Example 2.)

The amino acid sequence information generated in Example 1, together with published amino acid sequence information from the *Thermus aquaticus* DNA pol I gene, was used to create synthetic DNA primers for isolating a portion of the *Thermus flavus* DNA polymerase I gene. (Example  
10 3.) More particularly, a first primer, designated FTFL2, was synthesized to correspond with known coding sequence from *T. aquaticus* DNA pol I gene (Lawyer *et al.*, *J. Biol. Chem* 264: 6427-6437 (1989)), and to bind to the top strand of the *T. aquaticus* DNA pol I gene. The particular *T. aquaticus* coding sequence chosen encodes a portion of the *T. aquaticus* DNA pol I  
15 amino acid sequence that is homologous to the native *T. flavus* DNA pol I peptide that had previously been sequenced (Example 1). A second primer, designated RTFL4, was synthesized to have a sequence that binds to the 3'-end of the *T. aquaticus* gene on the opposite strand. A DNA amplification reaction was performed with primer FTFL2, primer RTFL4, and *T. flavus*  
20 genomic DNA. The amplification reaction yielded a single amplification product, designated the "2-4 fragment." This fragment was cloned into M13mp18 vector, amplified in *E. coli*, and sequenced.

As explained in detail in Example 4, the 2-4 fragment (obtained by the procedures outlined in Example 3) was used to isolate the *Thermus*  
25 *flavus* DNA pol I gene from the *T. flavus* genomic library that had been constructed (Example 2). Specifically, the 2-4 fragment was further amplified and used to generate probes via thermal cycle labeling (TCL). The amplified *T. flavus* genomic library was plated on 2XTY plates and grown until plaques formed. Duplicate plaque lifts were obtained from each plate onto Hybond

- 15 -

N filters, and these filters were then screened using the above-described TCL probes using hybridization methods well known in the art. Positive plaques were selected, purified by dilution and by re-screening with the 2-4 probes, and then further characterized. In particular, two clones with inserts of 14-16 kb, designated  $\lambda$ 21 and  $\lambda$ 51, were chosen for further analyses.

Clones  $\lambda$ 21 and  $\lambda$ 51 were used as a starting point from which the complete *T. flavus* DNA pol I gene was cloned and sequenced. As explained in detail in Example 5 and with reference to FIGURE 1A, restriction mapping, subcloning, and partial sequencing led to the determination that a subclone of  $\lambda$ 21 designated p21E10 contained about 2/3 of the Tfl DNA pol I gene (3' end), whereas a subclone from  $\lambda$ 51 designated p51E9 contained a 5' portion of the gene that overlapped the coding sequence contained in clone p21E10.

A primer walking procedure was used to obtain the complete sequence of the gene. Specifically, primers homologous or complimentary to the ends of previously determined sequences (obtained from p21E10 and from other deletion vectors) were synthesized and used in additional sequencing reactions. By repeating this process the entire length of the gene was sequentially sequenced. The DNA and deduced amino acid sequence for the *T. flavus* DNA pol I holoenzyme are shown in FIGURE 2, which corresponds to SEQ. ID NO: 1 and 2 in the Sequence Listing. The sequences of each primer used, and the relative location of the primers in the gene sequence, are depicted in Table 2 and in FIGURE 4, respectively. The amino acid sequence of the holoenzyme depicted in FIGURE 2 and SEQ. ID NO: 2 corresponds with nucleotides 301 to 2802 of the DNA depicted in FIGURE 2 and SEQ ID NO: 1.

The foregoing results demonstrate that an aspect of the invention is directed to a purified DNA encoding a thermostable polypeptide having DNA polymerase activity, the DNA comprising nucleotides 301 to 2802 of SEQ ID NO: 1. This DNA may be operatively linked to other

- 16 -

DNAs, such as expression vectors known in the art. The invention is also directed to a vector having at least one insert consisting essentially of nucleotides 301 to 2802 of SEQ ID NO: 1, the nucleotides encoding a thermostable polypeptide having DNA polymerase activity. Similarly, the invention is directed to a vector comprising nucleotides 301 to 2802 of SEQ ID NO:1, the nucleotides encoding a polypeptide having thermostable DNA polymerase activity.

With the gene sequence established, the DNA and deduced amino acid sequences of the *T. flavus* DNA pol I gene were aligned and compared to the DNA and deduced amino acid sequences of the purported Tfl DNA pol I published by Akhmetzjanov and Vakhitov, *Nucleic Acids Res.* 20:5839 (1992) (83% DNA sequence homology, 85% amino acid sequence homology) and to the deduced amino acid sequence of the Taq pol I gene (86% DNA sequence homology, 87% amino acid sequence homology). The amino acid comparison is depicted in FIGURE 3.

To produce a recombinant *T. flavus* DNA pol I protein a full-length *T. flavus* DNA pol I gene clone was constructed, expressed in *E. coli*, and purified. As detailed in Example 6 and FIGURE 1A, plasmids p51E9 and p21E10 were further restriction mapped and subsequently subcloned to generate plasmid p21BRV2, containing a 1.3 kb insert that includes the 3' region of the Tfl DNA pol I gene, and plasmid p51X16, containing a 2.5 kb BamHI fragment in which the 5' region of the gene was located. Linearization of plasmid p21BVR2 with BamHI and ligation of this linearized plasmid to the BamHI fragment of p51X16 yielded clone pTFL 1.4, containing the entire Tfl DNA pol I gene.

*E. coli* DH5 $\alpha$ F' were transformed with plasmid pTFL 1.4 and grown in a fermentor to recombinantly produce *T. flavus* DNA pol I holoenzyme. As detailed in Example 6, this recombinant protein was purified from the lysed *E. coli* with a method that included a heat denaturation of *E.*

- 17 -

*coli* proteins, precipitations and centrifugations, Sephadex G-25 and Bio-Rex 70 column chromatography, and immunoaffinity chromatography. The calculated DNA polymerase specific activity of *T. flavus* DNA pol I isolated by this procedure was determined to be 79,500 U/mg protein.

5 In order to increase the yield of recombinant *T. flavus* DNA pol I holoenzyme, a second expression clone was constructed in which the lacZ promoter was fused directly to the initiation codon of the Tfl DNA pol I gene. As detailed in Example 7 and FIGURE 1A, the promoter was fused to the 5' portion of the gene located using site-directed mutagenesis, and a second  
10 generation expression clone, designated pTFLRT4, was generated.

*E. coli* (strain DH5 $\alpha$ F'IQ) were transformed with pTFLRT4 and cultivated, and recombinant *T. flavus* DNA pol I was isolated therefrom and purified. As detailed in Example 7, the purification protocol includes heat treatment, polyethyleneimine- (PEI-) precipitation, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>- precipitation,  
15 Bio Rex 70 chromatography and immunoaffinity chromatography. The yield was approximately 2,000,000 units of enzyme from 500 g of cells, and the purified enzyme preparation was found to have a DNA polymerase specific activity of 217,600 U/mg protein. The N-terminal amino acid sequence of the recombinant Tfl DNA pol I enzyme was determined and found to be identical  
20 to the sequence deduced from the *T. flavus* DNA Pol I gene sequence.

The foregoing discussion demonstrates that an aspect of the invention is directed to a purified DNA comprising a portion of the insert of plasmid pTFLRT4, the plasmid having ATCC Accession No. 69633, the portion encoding a thermostable polypeptide having DNA polymerase activity.  
25 This DNA may be operatively linked to additional DNAs, such as promoter DNAs and/or expression vector DNAs known in the art. A preferred DNA is plasmid pTFLRT4 itself. The present invention is also directed to thermostable polypeptides having DNA polymerase activity. In one aspect, the invention is directed to a *Thermus flavus* DNA polymerase protein

- 18 -

substantially free of other *Thermus flavus* proteins. Exemplary proteins include a DNA polymerase protein having the amino acid sequence of SEQ ID NO: 2. Similarly, the invention is directed to a thermostable polypeptide having DNA polymerase activity and consisting essentially of the amino acid sequence of SEQ ID NO: 2.

In addition to the cloning and expression of the Tfl DNA pol I holoenzyme, a vector allowing for the expression of a truncated DNA polymerase was generated. As explained in Example 8 and FIGURE 1B, a vector lacking the 5' one-third of the *T. flavus* DNA polymerase I gene was constructed. Specifically, the ATG start codon of lacZ was brought in frame with the DNA encoding amino acids 239 to 834 of the Tfl DNA pol I holoenzyme using site-directed mutagenesis, and the resulting plasmid, designated p21EHcM1.1, was expressed in *E. coli* DH5 $\alpha$ F'. The insert of plasmid p21EHcM1.1 includes a DNA sequence that corresponds with SEQ ID NO: 3 in the Sequence Listing, and encodes a polypeptide predicted to have the amino acid sequence depicted in SEQ ID NO: 4. The expressed polypeptide product was designated *Thermus flavus* DNA polymerase I exonuclease-free fragment, or "Tfl exo' fragment." An aspect of the invention is directed to a purified DNA comprising a portion of the insert of plasmid p21EHcM1.1, the plasmid having ATCC Accession No. 69632, the portion encoding a thermostable polypeptide having DNA polymerase activity. This DNA may be operatively linked to additional DNAs, such as known promoter DNAs and/or expression vectors. A preferred DNA is plasmid p21EHcM1.1 itself.

As detailed in Example 8, the purification protocol for the Tfl exo' fragment expressed in *E. coli* [p20EHcM1-1] included PEI-precipitation, gel filtration, Procion-Red Sepharose chromatography and immunoaffinity chromatography. The yield using this preparation protocol was approximately



- 19 -

300,000 units of enzyme from 50 g of cells, and the preparation had a DNA polymerase specific activity of 600,000 U/mg protein.

The N-terminal amino acid sequence of the Tfl exo' fragment was determined (Example 8), and interestingly, the purified protein lacked 37 N-terminal amino acids predicted from the DNA encoding the exo' fragment. The deduced amino acid sequence of the purified Tfl exo' fragment -- based on this amino acid sequence data and the complete DNA sequence -- is depicted in SEQ ID NO: 5, and corresponds with amino acid 275 to 834 of FIGURE 2. A minor sequence lacking 38 N-terminal amino acids was also detected.

The foregoing demonstrates that another aspect of the invention is directed to a purified DNA encoding a thermostable polypeptide having DNA polymerase activity, the DNA comprising a portion of SEQ ID NO: 3. For example, the invention is directed to a purified DNA comprising nucleotides 112 to 1791 of SEQ ID NO: 3. This DNA also may be operatively linked to other DNAs, such as to nucleotides 1 to 111 of SEQ ID NO: 3, and/or to expression vectors known in the art. In a related aspect, the invention is directed to a vector comprising nucleotides 112 to 1791 of SEQ ID NO: 3, the nucleotides encoding a polypeptide having thermostable DNA polymerase activity. Similarly, the invention is directed to a vector having at least one insert consisting essentially of nucleotides 112 to 1791 of SEQ ID NO: 3, the nucleotides encoding a thermostable polypeptide having DNA polymerase activity.

The recombinant expression and purification of biologically active Tfl exo' fragment demonstrates additional aspects of the present invention. For example the present invention is directed to a purified fragment of *Thermus flavus* DNA polymerase I protein, the fragment having DNA polymerase activity. Exemplary fragments include a fragment having an amino acid sequence comprising amino acids 2 to 560 or 1 to 560 of SEQ. ID NO: 5, and a fragment encoded by the insert of plasmid

- 20 -

p21EHcM1.1, having ATCC Accession no. 69632. Also, the invention is directed to a polypeptide having DNA polymerase activity and consisting essentially of the amino acid sequence of SEQ ID NO: 5.

The foregoing description of methods and recombinant cells demonstrates that the present invention is directed to more than DNA's and polypeptides. Another important aspect of the invention is directed to a host cell transformed with a DNA, vector, or plasmid of the present invention, including those specifically mentioned above. Preferably, the host cell transformed with a DNA is capable of expressing a thermostable polypeptide encoded by the DNA, the polypeptide having DNA polymerase activity. By host cell is meant both prokaryotic host cells, including *E. coli* cells, and eukaryotic host cells.

In addition to being directed to DNAs, transformed cells, and polypeptides, the present invention is directed to various methods for using DNAs and polypeptides. For example, the invention is directed to a method for purifying a thermostable polypeptide having DNA polymerase activity comprising the steps of: transforming a host cell with a DNA to create a transformed host cell, the DNA selected from the DNA's of the present invention; cultivating the transformed host cell under conditions to promote expression of a thermostable polypeptide encoded by the DNA, the polypeptide having DNA polymerase activity; and purifying the thermostable polypeptide with a monoclonal antibody that is cross-reactive with the thermostable polypeptide. In one preferred method, the cross-reactive monoclonal antibody has specificity for a *Thermus aquaticus* DNA polymerase and/or for a *Thermus flavus* DNA polymerase.

In another preferred method, commercially available chromatography columns are used to purify the expressed polypeptide.

The purification protocols for recombinant Tfl DNA polymerase I and Tfl exo' fragment demonstrate that another aspect of the invention relates to methods of purifying a thermostable polypeptide having DNA polymerase

- 21 -

activity. One such method includes the steps of expressing the thermostable polypeptide in a host cell, the polypeptide having an amino acid sequence encoded by a DNA of the present invention; lysing the cell to create a suspension containing the thermostable polypeptide and host cell proteins and cell debris; contacting a soluble portion of the suspension with an antibody that is immunologically cross-reactive with the thermostable polypeptide under conditions wherein the antibody binds to the thermostable polypeptide to form an antibody-polypeptide complex; isolating the antibody-polypeptide complex; and separating the thermostable polypeptide from the isolated antibody-polypeptide complex to provide a purified thermostable polypeptide. Preferably, such a method further includes the steps of heating the suspension to denature the host cell proteins; and centrifuging the suspension to remove the cell debris and denatured host cell proteins. In more preferred methods, the immunologically cross-reactive antibody is a monoclonal antibody, such as a monoclonal antibody that is specific for *Thermus aquaticus* DNA polymerase I and/or *Thermus flavus* DNA polymerase I. This preferred method is exemplified herein using a monoclonal antibody purified from a hybridoma designated hybridoma 7B12. This monoclonal antibody is commercially available from Molecular Biology Resources, Inc., Milwaukee, Wisconsin, as Cat. No. 4100-01.

The invention is also directed toward a method of using a DNA encoding a DNA polymerase enzyme to generate active fragments of the DNA polymerase enzyme, comprising the steps of: deleting a portion of the DNA to create a modified DNA, expressing the modified DNA to produce a DNA polymerase enzyme fragment, purifying the DNA polymerase enzyme fragment, assaying the DNA polymerase enzyme fragment for DNA polymerase activity, and selecting a DNA polymerase enzyme fragment having DNA polymerase activity, wherein the DNA is selected from DNAs of the present invention.

- 22 -

As detailed in Example 9 and summarized in Table 3A, a number of experiments were conducted to characterize the exonuclease activities of *T. flavus* DNA pol I holoenzyme and *exo*<sup>-</sup> fragment. For both the holoenzyme and the *exo*<sup>-</sup> preparation, each exonuclease and endonuclease activity assayed was either very low or undetectable.

As detailed in Example 10, a number of additional assays were performed to better characterize the recombinant Tfl DNA pol I proteins that had been purified and to compare these proteins to other known thermostable DNA polymerases. For example, the DNA polymerase activity of the Tfl holoenzyme and the *exo*<sup>-</sup> fragment was analyzed at different pH values, and at different MgCl<sub>2</sub> and MnCl<sub>2</sub> concentrations. FIGURES 5 (pH optima); 6A (MgCl<sub>2</sub> optima); 6B (MnCl<sub>2</sub> optima) and 7A (temperature optima) summarize the results of some of these assays. The optimal range and the peak values (in parentheses) are summarized in Table 1A.

15

TABLE 1A		
	Holoenzyme	Exo <sup>-</sup> Fragment
pH	9.5 - 10.5 (10)	7.5 - 10 (8.5)
MgCl <sub>2</sub> [mM]	>50	1.3 - 13 (5)
MnCl <sub>2</sub> [mM]	0.8 - 4 (2)	2.1 - 11 (4)

20

To assay thermostability the enzymes were incubated for 30 min. at different temperatures to define the temperature optimum. The highest activity (100%) was found at 80°C for the holoenzyme (14% remaining after 30 min. at 90°C), and 70 to 75°C (8% remaining after 30 min. at 90°C) for the *exo*<sup>-</sup> fragment.

25

The Tfl holoenzyme preparation enzyme was more than 95% pure as judged by sodium dodecyl sulfate polyacrylamide gel electrophoresis

- 23 -

(SDS-PAGE) on 12.5% gel (FIGURE 9). The apparent molecular weight was 80 kD, which is lower than the calculated molecular weight of approximately 94 kD based on the DNA sequence. The Tfl holoenzyme preparation was found to be free of detectable double-stranded nucleases and of 5' → 3' exonuclease and endonuclease activities. Low levels of single-stranded nucleases and of 3' → 5' exonuclease activity were found. The isoelectric point was determined to be 6.43.

The purified Tfl exo' fragment was found to possess low 3' → 5' and 5' → 3' exonuclease activities. The preparation was more than 95% pure as judged by SDS-PAGE (FIGURE 9). The apparent molecular weight of 68 kD as judged by SDS-PAGE compares well with the calculated molecular weight of approximately 63 kD. The Tfl exo' preparation was found to be free of detectable double- and single-stranded nucleases and endonuclease activities. The isoelectric point was determined to be 5.94.

The performances of the Tfl holoenzyme and the exo' fragment were tested in DNA sequencing, PCR and TCL (Examples 11, 12 and 13). Both enzymes were found to be useful in sequencing reactions utilizing labeled primer in conjunction with single-stranded and double-stranded DNA templates, and in a cycle sequencing reaction with a single-stranded template. The enzymes were also useful in sequencing reactions utilizing internal labeling with, for example, [ $\alpha^{35}\text{S}$ ]-dATP. In all the reactions tested the Tfl exo' fragment provided DNA sequence information of more than 150 nucleotides, as did recombinant Tfl DNA pol I holoenzyme.

The Tfl DNA pol I holoenzyme and the exo' fragment were tested in PCR reactions. The recombinant holoenzyme gave similar results to the native enzyme. The Tfl exo' fragment retained 50% of its activity after 16 cycles. The holoenzyme retained 50% of its activity for 20 cycles. The specific amplified products were analyzed at the same time. After 20 cycles, an amplification product was visible on agarose gels. The amount of product increased between 25-50 cycles, but decreased after 100 cycles.

- 24 -

The native *T. flavus* enzyme provided with the ZEPTO™ Labeling kit (CHIMERx, Madison, WI) was replaced by the recombinant holoenzyme or by the (recombinant) *exo'* fragment. The efficiency of the labeling of plasmid pUC19 was determined on agarose gels and the efficiency of incorporation was determined in dot blot analysis. A dilution of  $1:10^6$  of labeled probes generated with the holoenzyme was detectable ( $1:10^5$  for probes generated by the *exo'* fragment). Both results indicated that the enzymes have the required activity needed for labeling pUC19 DNA in TCL.

A protocol is also provided for demonstrating that the present invention is also directed to TCL in which the recombinant Tfl DNA pol I holoenzyme is employed without exogenous primers for enzymatic extension. In this method, referred to as Universal Thermal Cycle Labeling (UTCL), DNA of unknown sequence is combined intact with rTfl DNA Pol I holoenzyme, deoxyribonucleotide triphosphates, and the appropriate buffer. The holoenzyme is then combined with intact template and subjected to repeated cycles of denaturation annealing and extension. Alpha  $^{32}\text{P}$ -dATP,  $^{32}\text{P}$ -dTTP,  $^{32}\text{P}$ -dGTP,  $^{32}\text{P}$ -dCTP, biotin-dUTP, fluorescein-dUTP, or digoxigenin-dUTP is also included in the extension step for subsequent detection purposes.

The foregoing results demonstrate further aspects of the invention. For example, the invention is further directed to a method for labeling DNA, comprising the steps of: digesting an aliquot of template DNA with a restriction endonuclease reagent wherein the digestion generates sequence-specific DNA fragments; mixing an aliquot of undigested template DNA with the sequence-specific DNA fragments; denaturing the mixture of template DNA and sequence-specific DNA fragments thereby generating denatured template DNA and oligonucleotide primers; annealing the primers to the denatured undigested template DNA to form a DNA-primer complex; and performing an extension reaction from the primers in the DNA-primer

- 25 -

complex using Tfl *exo*<sup>-</sup> fragment in the presence of one or more nucleotide triphosphates, wherein at least one nucleotide triphosphate has a label.

Further, the invention is directed to a method for thermal cycle labeling DNA comprising the steps of: digesting an aliquot of template DNA with a restriction endonuclease reagent wherein the digestion generates sequence-specific DNA fragments; mixing an aliquot of undigested template DNA with the sequence-specific DNA fragments; denaturing the mixture of template DNA and the DNA fragments thereby generating denatured template DNA and oligonucleotide primers; annealing the primers to the denatured undigested template DNA to form a DNA-primer complex; performing an extension reaction from the primers in the DNA-primer complex using Tfl DNA pol I *exo*<sup>-</sup> fragment in the presence of one or more nucleotide triphosphates wherein at least one nucleotide triphosphate has a label; heat-denaturing the labeled extension products; reannealing the excess primers with the template DNA and with the extension products; and performing at least one additional extension reaction from the DNA-primer complex using a Tfl DNA pol I *exo*<sup>-</sup> fragment.

The present invention is further directed to kits for labeling DNA. A kit of the present invention includes, in association: a labeling buffer; a concentrated mixture of 1 or more nucleotide triphosphates; Tfl DNA pol I *exo*<sup>-</sup> fragment; and a control DNA, the control DNA being useful for monitoring the efficiency of labeling. Additionally, the kit may include a restriction endonuclease reagent and a restriction endonuclease buffer.

In another aspect, a kit of the present invention for labeling DNA comprises, in association: a Tfl DNA pol I *exo*<sup>-</sup> fragment; and a Tfl DNA pol I *exo*<sup>-</sup> fragment buffer. Preferably, such a kit further comprises a concentrated mixture of 1 or more nucleotide triphosphates and a control DNA, the control DNA being useful for monitoring the efficiency of labeling.

The following examples are intended to describe various aspects of the invention in greater detail. More particularly, in Example 1, the

purification and amino acid sequencing of native *Thermus flavus* DNA polymerase I is described. In Example 2, the construction and amplification of a *Thermus flavus* genomic DNA library is described. In Example 3, the cloning and sequencing of a *Thermus flavus* DNA polymerase I gene fragment is described. Example 4 details the preparation of gene-specific probes and screening of the *Thermus flavus* genomic library for clones containing the *T. flavus* DNA pol I gene. Example 5 details the sequencing of the *T. flavus* DNA polymerase I gene. In Example 6, the construction and expression of a full-length *T. flavus* DNA pol I clone and purification of full-length recombinant *T. flavus* DNA pol I protein are described. In Example 7, the construction and expression of a high-yield, full-length *T. flavus* DNA pol I clone and purification of full-length recombinant *T. flavus* DNA pol I is described. Example 8 details the cloning and expression of the exo' fragment of *T. flavus* DNA polymerase I. In Example 9, the characterization of recombinant *T. flavus* DNA polymerase I exonuclease activities is detailed. In Example 10, studies are described comparing the recombinant *T. flavus* and *T. aquaticus* DNA polymerases. In Example 11, DNA sequencing with recombinant *T. flavus* DNA polymerases is detailed. Example 12 demonstrates the utility of recombinant Tfl holoenzyme and the exo' fragment in polymerase chain reaction procedures. Example 13 demonstrates the utility of recombinant Tfl DNA pol I holoenzyme and the Tfl exo' fragment for use in thermal cycle labeling procedures. Example 14 analyzes the utility of *T. flavus* DNA pol I holoenzyme and exo' fragment for reverse transcription applications. Example 15 demonstrates the increased processivity of Tfl exo' fragment as compared to native or recombinant Tfl DNA pol I holoenzyme or Taq holoenzyme. Finally, Example 16 details a large "production scale" purification of recombinant Tfl holoenzyme and exo' fragment.



- 27 -

## EXAMPLE 1

Purification and Amino Acid  
Sequencing of Native Tfi DNA Pol I

Native *T. flavus* DNA polymerase I was isolated from *T. flavus* cells and used to generate amino acid sequence information as described below.

*Thermus flavus* obtained from the American Type Culture Collection (ATCC 33923, Catalogue of Bacteria and Bacteriophages, 18th Edition, 1992) was cultured as follows: one ampule of *Thermus flavus* ATCC 33923 was used to inoculate 100 ml culture medium (0.1 g nitrilotriacetic acid, 3 g NZ Amine A, 3 g yeast extract, 5 g succinic acid [free acid], 0.001 g riboflavin, 0.522 g  $K_2HPO_4$ , 0.480 g  $MgSO_4$ , 0.020 g NaCl, 2 ml Trace Metal Solution (0.5 ml  $H_2SO_4$ , 2.2 g  $MnSO_4$ , 0.5 g  $ZnSO_4$ , 0.5 g  $H_3BO_3$ , 0.016 g  $CuSO_4$ , 0.025 g  $Na_2MoO_4$ , 0.046 g cobalt nitrate) per liter, adjusted to pH 8.0 with NaOH) and the culture was incubated overnight at 70°C with shaking. In the morning 10 ml of the overnight culture was used to inoculate 1000 ml of medium. This culture was grown for about 8 hours at 70°C and then used as an inoculum for 170 liters of medium in a New Brunswick 250 liter fermentor equipped with a ML 4100 controller. The settings for a typical fermentation were 3 pounds back pressure, 60 liters/min. (lpm) aeration, 100 rpm agitation, at 70°C. The fermentation was terminated when the cells reached a density of 2 - 3 O.D., as measured at 600nm. The cells were cooled down to room temperature and harvested by centrifugation at 17,000 rpm in a CEPA type 61 continuous flow centrifuge with a flow rate of 2 lpm. The cell paste was stored at -70°C until used.

*T. flavus* cells (500-1500g) were thawed in 3 volumes of lysis buffer (20mM Tris-HCl, pH 8.0, 0.5mM ethylenediaminetetraacetate (EDTA), 7 mM  $\beta$ -mercaptoethanol ( $\beta$ ME), 10 mM  $MgCl_2$ ) and homogenized. Phenylmethylsulfonyl fluoride (PMSF), a protease inhibitor, was added to a final concentration of 0.3 mM. The suspension was then treated with 0.2 g/l

- 28 -

of lysozyme (predissolved in lysis buffer) at 4°C for 1 hr. Cells were homogenized twice at 9000 psi in a Manton Gaulin homogenizer, with the suspension chilled to approximately 10°C between passes. New PMSF was added to 0.2 g/l before, between and after passes. NaCl and polyethyleneimine (PEI) (10% w/v, pH 7.0) were added to the crude, homogenized lysate to a final concentration of 0.5 M and to 0.2%, respectively. The sample was mixed well and centrifuged at 13,500 x g for 1 hour.

The supernatant from the centrifuged lysate was desalted by diluting with 10 liters of DE52 column buffer (20 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, 7 mM  $\beta$ ME) and concentrated to approximately 4 liters using an Amicon S10Y30 Spiral Ultrafiltration cartridge. The dilution/concentration step was repeated two times, with a final concentrated volume of about 4 liters.

The desalted sample was batch contacted with 400 g of equilibrated Whatman DE52 ion exchange resin (Maidstone, England). The suspension was collected on a sintered glass funnel and washed 3 times with 1 volume of DE52 column buffer. The resin was then resuspended in a minimal volume of buffer and poured into a column (4.5 x 50 cm), packed and washed with an additional volume of buffer. The column was eluted with a 0-0.5 M NaCl linear gradient (total gradient volume: 2000 ml). Twenty-five ml fractions were collected at a rate of about 5 ml/min. Peak fractions (fractions containing DNA polymerase activity) were determined by a modified DNA polymerase assay described by Kaledin et al., *Biokhimiya* 45:644-651 (1980), pooled and dialyzed in approximately twenty-five volumes of Affi-Gel Blue (AGB) column buffer (20 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 10 mM  $\beta$ ME, 10mM  $MgCl_2$ , 0.02% Brij 35).

The dialyzed DE52 peak fractions were applied to an AGB column (4.4 x 40 cm, 600 ml packed volume, MBR Blue, Molecular Biology Resources, Milwaukee, WI), which was washed with 2 column volumes of

- 29 -

AGB column buffer, and eluted with a 0-1.2 M NaCl linear gradient (total gradient volume: 2000 ml). Twenty-five ml fractions were collected at a rate of 1-5 ml/min. The peak fractions were dialyzed as above in AGB buffer.

- 5 The dialyzed AGB peak fractions were applied to a heparin agarose column (4.4 x 16.5 cm, 250 ml packed volume (Affigel Heparin, Bio-Rad, Hercules, CA; or Heparin Agarose, Molecular Chimetrics, Madison, WI)), which was washed with approximately 2 column volumes (until effluent is no longer colored, and column resin is white in appearance), and eluted with a 0.1-1.0 M NaCl linear gradient (total gradient volume: 1500 ml).
- 10 Twenty-five ml fractions were collected at a rate of 1-5 ml/min. The peak fractions were dialyzed in HP Q Sepharose Column Buffer (20 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 7 mM  $\beta$ ME, 0.1% Brij 35).

- The dialyzed heparin agarose peak fractions were filtered through a 0.2  $\mu$ m filter and applied at 4 ml/min. to the HP Q Sepharose column (Pharmacia, Uppsala, Sweden) on FPLC. The column was washed with several column volumes of buffer, and eluted with a 0-0.25 M NaCl linear gradient. Ten ml fractions were collected at 4 ml/minute. The peak fractions were dialyzed in HP S Column Buffer (20 mM Na-Citrate, pH 6.0, 1 mM EDTA, 7 mM  $\beta$ ME, 0.1% Brij 35) or diluted in the same buffer,
- 20 depending on the volume of the fraction pool.

- The dialyzed (or diluted) HP Q peak fractions were filtered through a 0.2  $\mu$ m filter and the HP S column (Pharmacia) was run as above, washing with HP S Column buffer and eluting with a 0-0.25 M NaCl gradient. Peak fractions were pooled and dialyzed against 4 liters of Final
- 25 Storage Buffer (50 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 5mM DTT, 50% glycerol). The final product was diluted to a concentration of 5000 U/ml in the above buffer including 0.5 % Tween 20 (Sigma Chemical Co., St. Louis, MO) and 0.5 % Nonidet P40 (Fluka Biochemika, Buchs, Switzerland) as stabilizers and stored at -20°C. A typical preparation from 1200 g of cells

- 30 -

yields approx. 2,000,000 units (1,700 units/g) or about 40 mg of DNA polymerase.

- To quantify DNA polymerase activity, a DNA polymerase activity assay was performed using a modification of a protocol described by Kaledin et al., *Biokhimiya* 45:644-651 (1980). Reactions were performed in a 50  $\mu$ l reaction mixture of 25 mM Tris-HCl, pH 9.5 at 23°C; 50mM KCl; 10 mM MgCl<sub>2</sub>; 1 mM DTT; 0.2 mM each dCTP, dGTP, dTTP, pH 7.0; 0.2 mM [ $\alpha$ -<sup>32</sup>P]dATP, pH 7.0, 10  $\mu$ Ci/ml; 50  $\mu$ g BSA; 15  $\mu$ g activated DNA (Baril et al. *Nucleic Acids Res.* 8:2641-2653 (1977)); and 5  $\mu$ l of diluted enzyme. For control purposes enzymes (in general AmpliTaq DNA polymerase (Perkin Elmer, Cat. No. N801-0060), or Taq DNA polymerase purified according to a procedure described by Kaledin et al., *Biokhimiya* 45:644-651 (1980)) with known activities are diluted to 20, 40 and 80 units/ml. Two reactions were run without enzyme as negative controls for background subtraction.

- A 45  $\mu$ l reaction mixture, less enzyme, was prepared and the reaction was started by the addition of 5  $\mu$ l of enzyme. After 10 min. of incubation at 70°C, 40  $\mu$ l was removed and added to 50  $\mu$ l of yeast RNA co-precipitant (10 mg/ml in 0.1 M sodium acetate, pH 5.0). One ml of 10% trichloroacetic acid (TCA) was added and the samples were placed on ice for at least 10 minutes. The mixture was filtered on a glass fiber filter disc and washed first with 5% TCA/ 2% sodium pyrophosphate, and then with 95% ethanol. The dried filter disc was counted in 5 ml of scintillation fluid.

- One unit of activity is defined as the amount of enzyme required to incorporate 10 nmol of total nucleotide into acid insoluble form in 30 min. at 70°C in this assay, the standard activity assay.

To estimate protein concentration, an aliquot of a native *T. flavus* DNA polymerase preparation (1100 U/ml) was separated on a 5 - 25% SDS-polyacrylamide gel, using the Bio-Rad protocols (Hercules, CA) and the

- 31 -

Bio-Rad Mini-Protean II electrophoresis unit. The concentration was estimated at 33  $\mu\text{g}/\text{ml}$  when compared to co-electrophoresed protein standards.

To obtain amino acid sequence information from native *T. flavus* DNA polymerase, about 53  $\mu\text{g}$  of native polymerase were separated on a preparative 7.5% SDS-polyacrylamide gel, blotted onto PVDF membrane and stained with amido black as described by Matsudaira, *J. Biol. Chem.* 262: 10035-10038 (1987). The major band at approximately 83 kD was excised and sequenced using an Applied Biosystems (Foster City, CA) 477A Protein Sequencer. No N-terminal sequence was obtained under these conditions.

Due to the apparent block at the N-terminus of the native *T. flavus* DNA polymerase I (holoenzyme), another approach was employed to obtain a partial amino acid sequence. Native *T. flavus* DNA polymerase I was digested with trypsin, and the resulting peptides were separated using reverse phase high-performance liquid chromatography (HPLC). The N-terminal amino acid sequences of four of these peptides (peptides 1-4) were determined. The amino acid sequence of one of the peptides, peptide 1, is LHTRFNQTATATGRLSSDPNLQNIPVR. This sequence has been determined to map at positions 562 to 589 in the deduced amino acid sequence of the Tfl DNA pol I holoenzyme described herein (FIGURE 2)). As explained in Example 3, knowledge of this amino acid sequence information was used to isolate the *T. flavus* DNA polymerase I gene.

## EXAMPLE 2

### Construction and Amplification of a *Thermus flavus* genomic DNA library

A *Thermus flavus* genomic library was constructed in phage  $\lambda$  Dash II and amplified in the following manner.

Genomic DNA from the *Thermus flavus*, cultured overnight as described above, was isolated according to the procedure described by Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing

- 32 -

Associates and John Wiley & Sons, New York (1990). In general, yields of genomic DNA between 100 and 900  $\mu\text{g}$  were obtained from the cell pellet of about 1.5 ml of culture.

Twenty-five micrograms of *Thermus flavus* genomic DNA were partially digested with 0.3 units of Sau3A1 in a total reaction volume of 50  $\mu\text{l}$ . At 0, 5, 10, 15, and 30 min., 10  $\mu\text{l}$  samples were removed and the enzyme was inactivated at 65°C for 15 min. An aliquot from each time point was analyzed on a 1.2% agarose/TBE gel. The 10 min. reaction time produced fragments having the desired size distribution (3 kb to 20 kb).

Approximately 2.5 pmoles of 5'-ends of Sau3A1-digested *T. flavus* DNA were treated with calf intestinal alkaline phosphatase (CIP) using standard techniques (Ausubel *et al.*, Current Protocols in Molecular Biology (1990)). One half of the CIP-digested sample was electrophoresed on a 0.7% agarose gel and checked for amount and integrity. The Sau3A1- digested, CIP-treated *T. flavus* DNA was extracted with phenol/chloroform and chloroform, ethanol precipitated, pelleted, and washed in 70% ethanol. The pellet was stored at -20°C. This DNA is referred to as "CIP TFL DNA."

The *T. flavus* library was constructed as described in the manufacturer's instructions using the phage  $\lambda$  DASH II / BamHI Cloning Kit (Stratagene, LaJolla, CA) and the CIP Tfl DNA. The pME/BamHI test insert (0.3  $\mu\text{g}$ ) was run in parallel as a control. The ligation mixture was incubated over night at 4°C.

The *T. flavus* DNA ligated to  $\lambda$  DASH II arms was packaged *in vitro* using the Gigapack II Gold Packaging Extract from Stratagene, according to the manufacturer's instructions. Control DNA provided by the manufacturer was also packaged.

Following the protocol provided by Stratagene with the  $\lambda$  DASH II / BamHI Vector Kit, host bacteria were prepared: *Escherichia coli* VCS 257 (Stratagene) for wild type phage; *E. coli* SRB and SRB(P2)

- 33 -

(Stratagene) for the *T. flavus* library and the control. VCS 257 was grown in NZY + maltose medium; SRB and SRB(P2) were grown in NZY + maltose medium with 50 µg/ml kanamycin at 37°C for 6 hours. After centrifugation of the cells at 2800 x g for 10 min., the cells were resuspended in sterile 10 mM MgSO<sub>4</sub> to give an A<sub>600</sub> (optical density at 600nm) of 0.5.

Two 1:10 serial dilutions were prepared from the control phage and the CIP Tfl DNA library. Ten microliters of undiluted, 1:10, and 1:100 dilutions of phage were added to 200 µl of SRB cells. The cells were incubated with light shaking at 37°C for 15 minutes and after the addition of top agar, the mixture was poured onto LB/M/M plates. The plates were incubated overnight at 37°C.

The *T. flavus* library was amplified using techniques described by Ausubel *et al.*, Current Protocols in Molecular Biology (1990), and the primary and amplified libraries were titered on SRB cells and the titers are shown in Table 1B. The amplified library was stored at 4°C.

TABLE 1B		
Titer (plaque forming units/ml)		
Construct	Primary Library	Amplified Library
CIP TFL DNA	4.4 x 10 <sup>5</sup>	9.6 x 10 <sup>7</sup>
pME/BamHI	3 x 10 <sup>6</sup>	1.5 x 10 <sup>9</sup>
λ Control DNA	1.1 x 10 <sup>9</sup>	Not determined

### EXAMPLE 3

#### Cloning and Sequencing a Tfl DNA Pol I Gene Fragment

The amino acid sequence information derived from four Tfl DNA pol I peptides (Example 1) was used to design the synthesis of two primers for the amplification of a *T. flavus* DNA polymerase gene fragment:

- 34 -

primer FTFL2 (primer "2"; 21mer) (SEQ ID NO: 8) and primer RTFL4 (primer "4"; 25mer) (SEQ ID NO: 9) (synthesized by Synthetic Genetics, San Diego, CA). The sequence of the two primers was also compared to the *T. aquaticus* DNA polymerase sequence (Lawyer et al., *J. Biol. Chem* 264: 6427-6437 (1989)); the primer nucleotide sequences, with cross-references to the Sequence Listing and Sequence ID Nos. are shown in Table 2. Primer FTFL2 was chosen because the amino acid sequence obtained from peptide 1 (Example 1) was identical to a sequence in the Taq DNA polymerase I protein. Primer FTFL2 corresponds to nucleotides 1719-1740 of the *T. aquaticus* DNA polymerase coding sequence, top strand (i.e., to a portion of the sequence that encodes a portion of the Taq DNA pol I protein that is homologous to Peptide 1). Primer RTFL4 hybridizes to the 3'-end of the Taq DNA pol I gene at position 2476 - 2500 and has sequence identical to the bottom strand (Lawyer et al., *J. Biol. Chem* 264: 6427-6437 (1989)).

A typical amplification reaction (100  $\mu$ l) contained 0.2 mM deoxynucleotide triphosphates (dNTPs), 1 x Taq Polymerase Reaction Buffer (10 x buffer is 100 mM Tris-HCl, pH 8.4, 500 mM KCl, 15 mM MgCl<sub>2</sub>), 0.5  $\mu$ M of each primer FTFL2 and RTFL4 (primer set 2-4), 50  $\mu$ l mineral oil and 15 ng *T. flavus* genomic DNA. After the initial denaturation step (Step 1), 2.5 units of AmpliTaq DNA polymerase (Perkin Elmer No. N801-0060, Foster City, CA) were added. Negative control reactions containing either no enzyme or no template were performed. The amplification program was carried out in a thermocycler as follows: Step 1: 95°C for 5 min.; Step 2: hold at 72°C (for the time required to add the enzyme); Step 3: 55°C for 45 sec.; Step 4: 72°C for 5 min.; Step 5: 95°C for 15 sec.; Step 6: repeat Steps 3-5 34 times; Step 7: 55°C for 45 sec.; Step 8: 72°C for 20 min.; Step 9: hold at 4°C until processing the product. Under these conditions primer set 2-4 gave a single amplification product from *T. flavus* genomic DNA. The observed mobility of the amplification product ("the 2-4 fragment") in 1% and



- 35 -

1.2% agarose gels was in agreement with the 782 bp predicted from the *T. aquaticus* coding sequence.

The 2-4 fragment was cloned, sequenced, and compared to a previously published DNA sequence for a purported *T. flavus* DNA polymerase

5 I as follows.

First, to improve cloning efficiency, the 2-4 fragment was fractionated, blunt ended, and phosphorylated as follows. Approximately 20  $\mu$ l of the 2-4 fragment was loaded onto a Sephacryl S-500 (400  $\mu$ l in a spin filter, preswollen, pre-equilibrated and stored in 100 mM Tris-HCl, pH 8.0, 10 1 mM EDTA) column and centrifuged at 2,000 x g for 5 min. to trap the unused primers from the PCR reaction. DNA that passed through the column was ethanol-precipitated and resuspended in double-distilled water (ddH<sub>2</sub>O). The 2-4 fragment was blunt-ended using mung bean nuclease (MBN) (Molecular Biology Resources, Inc., Cat. No. 1190-01.), and phosphorylated 15 with T4 polynucleotide kinase (Molecular Biology Resources, Inc., Cat. No. 1260-01) prior to ligation to a vector by procedures well known in the art.

M13mp18 RF DNA (Life Technologies, Grand Island, NY) was restriction-digested with Hinc II and Ecl 136 II to create blunt ends for ligation to the above 2-4 fragment. The vector ends were dephosphorylated 20 with CIP to reduce the probability of self-ligation. The digested and dephosphorylated M13mp18 vector and the 2-4 fragment (fractionated/MBN/kinased) were ligated for 2 hr. at room temperature using procedures that are well known in the art.

Using standard techniques, 5  $\mu$ l of the ligation reaction were 25 added to 50  $\mu$ l of DH5 $\alpha$ F' *E. coli* (Life Technologies, Grand Island NY) cells, which had first been made competent, and the cells were transformed (see Ausubel et al., Current Protocols in Molecular Biology (1990)). Different numbers of cells were spread on 2XTY plates and grown until plaques appeared. Several plaques were picked, and DNA was prepared using 30 the Minute Miniprep ssDNA Purification Kit (CHIMERx, Madison WI). It

- 36 -

was determined that plaques designated M13-TFL 4.21 and 4.22 contained the 2-4 fragment in opposite orientations.

The DNAs M13-TFL 4.21 and 4.22 were sequenced by Sanger's dideoxy method (Sanger et al., *Proc. Natl. Acad. Sci.* 74:5463 (1977)) using the SEQUAL™ Sequencing Kit from CHIMERx. The forward sequencing primers (FSP, Table 2, SEQ ID NO: 6) used in sequencing M13-TFL 4.21 and M13-TFL 4.22 were end-labeled using [ $\gamma^{32}$ P]-ATP and T4 polynucleotide kinase. The extension/termination reactions were performed according to the protocol provided with the SEQUAL™ Sequencing Kit (CHIMERx). One microliter of each extension/termination reaction was loaded onto a 6% sequencing gel, which was electrophoresed at 3000 volts for 3 hours. The bands were detected by autoradiography and the sequence was determined.

When the nucleotide sequence from both ends of the 2-4 DNA was aligned, 771 bp of the approximately 780 bp DNA could be determined. This 2-4 DNA sequence was compared with the purported *T. flavus* DNA polymerase sequence reported by Akhmetzjanov and Vakhitov, *Nucleic Acids Res.* 20:5839 (1992), which theoretically should have been amplified by this primer set. Eighty-four percent maximum matching, as calculated by the MacDNAsis software program (Hitachi Software Engineering America, San Bruno, CA), was found. This degree of homology, compared with a reported DNA polymerase gene, suggested that the 2-4 DNA was indeed part of the *T. flavus* (ATCC 33923) gene and could serve as a useful probe for screening the *T. flavus* genomic library. The homology of 84% also suggested that either (1) the purported *T. flavus* strain studied by Akhmetzjanov and Vakhitov and the *T. flavus* strain (ATCC 33923) do not have identical DNA polymerase I genes; or (2) the two strains have more than one gene or gene-like sequences having homology to DNA polymerase I genes.

- 37 -

## EXAMPLE 4

**Preparation of Gene-Specific Probes and Screening  
of the *Thermus flavus* Genomic Library for Clones  
Containing the *T. flavus* DNA Pol I Gene**

5           The 2-4 fragment described in Example 3 was used to isolate  
the *Thermus flavus* DNA pol I gene from the *T. flavus* genomic library  
(Example 2). Using M13-TFL 4.21 as template and primers FTFL2 and  
RTFL4, the 2-4 fragment was amplified by PCR as described above to obtain  
10 larger quantities of the fragment for use in preparing probes to screen the *T.*  
*flavus* genomic library. The amplified 2-4 fragment, migrating at about 780  
bp, was cut out of preparative 0.7% agarose gel, eluted, phenol-chloroform  
extracted and ethanol-precipitated. Approximately 1  $\mu$ g of the 2-4 fragment  
was digested with CviJI\* (CHIMERx) to generate sequence specific primers  
15 for labeling. A variety of thermal cycle labeling (TCL) probes were prepared  
with the 2-4 intact fragment (i.e., Biotin-11-dUTP, fluorescein and  
[ $\alpha^{33}$ P]dCTP probes) in the manner described below. Each set of duplicate  
plaque lifts or targets was screened using two different types of labeled  
probes. Digestion with CviJI\*, as well as this method of labeling, is  
described in a co-owned, copending U.S. Patent Application Ser. No.  
20 08/217,459, filed March 24, 1994, entitled "Methods and Materials for  
Restriction Endonuclease Applications," incorporated herein by reference in  
its entirety. The PCT counterpart of this application, filed March 24, 1994,  
is PCT App. No. US94/03246.

25           The 2-4 intact fragment was labeled with Biotin-11-dUTP as  
described in the manual for the ZEPTO™ Labeling Kit (CHIMERx). To  
determine the relative efficiency of the amplification reaction, 5  $\mu$ l of the  
amplified 2-4 TCL probe was electrophoresed on a 0.7% agarose gel along  
side a 1 kb molecular size ladder. The amplified probe was evident as a  
smear from 0.1-5 kb, which is an indication of a successful TCL reaction.

- 38 -

To determine the efficiency of incorporation of the biotin-11-dUTP, a dot blot assay was performed as follows: A serial dilution of the probe from 1:10 to 1:10<sup>8</sup> was made in TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and 1  $\mu$ l of each dilution was spotted on a Hybond-N membrane (Amersham, Arlington Heights, IL), UV-cross-linked for 3 min., followed by colorimetric detection of the incorporated biotin-11-dUTP using streptavidin-alkaline phosphatase as described in the ZEPTO™ labeling manual. The probe was detected at 1:10<sup>6</sup> dilution suggesting that the biotin-labeled 2-4 fragment was efficiently labeled and is highly sensitive for the screening of the Tfl genomic library.

The fluorescein labeled 2-4 fragment was prepared and analyzed as above except fluorescein-12-dUTP was used instead of biotin-11-dUTP. The fluorescein-12-dUTP incorporation was detected using alkaline phosphatase conjugated anti-fluorescein antibody (Boehringer-Mannheim, Indianapolis, IN) instead of streptavidin-alkaline phosphatase. These probes were detected at a 1:10<sup>6</sup> dilution by the colorimetric assay as described in the ZEPTO™ labeling manual or by chemiluminescence. Both the biotin and fluorescein non-radioactive probes were aliquoted and used throughout the entire screening process.

The preferred detection method for both the biotin-11-dUTP probes and the fluorescein-12-dUTP probes was chemiluminescence. For this method of detection the filters with hybridized probes were incubated either with streptavidin-alkaline phosphatase or alkaline phosphatase conjugated to anti-fluorescein antibody for 30 min. at room temperature. They were then rinsed three times with wash buffer (1 x phosphate buffered saline (PBS), 0.3% Tween 20 (Sigma Chemical Co., St. Louis, MO) 0.02% Na-azide) for 15 min. each and finally in assay buffer (0.1 M diethanolamine, 1 mM MgCl<sub>2</sub> and 0.02% Na-azide, pH 10) for 5 minutes. They were finally incubated in assay buffer containing CSPD™ (Tropix, Bedford, MA) a chemiluminescence substrate, for 15 min. in the dark followed by exposure to X-ray films. The

- 39 -

normal exposure times for the biotin-11-dUTP probes were 5-30 min. and for the fluorescein-12-dUTP probes were 2-6 hours.

The 2-4 intact fragment was labeled with [ $\alpha^{33}\text{P}$ ]dCTP as described in the ZEPTO™ labeling manual; a total of  $6 \times 10^7$  cpm of [ $\alpha^{33}\text{P}$ ]dCTP at  $1 \times 10^9$  cpm/ $\mu\text{g}$  was incorporated. For probes,  $1.5 \times 10^6$  cpm of radio-labeled DNA was added to each plaque lift.

The sensitivity and specificity of the labeled probes was demonstrated by screening blots of digested *T. flavus* genomic DNA. Specifically, *T. flavus* genomic DNA was restricted with different restriction enzymes, such as BamHI, BglI, DraI, EcoRI, EcoRV and PacI. 250 ng/lane of restricted DNA, along with 500 ng of IL-3A viral DNA as negative control (Xia, Y., et al., *Nucleic Acids Research* 15: 6075-6090 (1987)), were electrophoresed on a 0.7% agarose gel. A Southern transfer of this gel onto Hybond-N was prepared. The denatured DNA on the Southern blots was UV-cross-linked to the filter for 3 minutes. Duplicate blots were prehybridized in 2 ml of hybridization buffer (50% deionized formamide, 7% SDS, 120 mM Na phosphate, pH 7.2, 250 mM NaCl, 1 mM EDTA and 1 mM cetyldimethylethylammonium bromide and 20  $\mu\text{l}$  of denatured salmon sperm DNA at 10 mg/ml) in a heat-sealed plastic bag at 52°C for 1 hour. Seven  $\mu\text{l}$  of either the biotin-11-dUTP 2-4 TCL probe or the fluorescein-12-dUTP 2-4 probe was added to one set of the Southern blots and  $1.5 \times 10^6$  cpm of the [ $\alpha^{33}\text{P}$ ]dCTP 2-4 TCL probe was added to the duplicate blot. The filters were hybridized by incubation overnight at 52°C.

The filters with the radioactive probe were incubated with low stringency buffer (1 x SSC, 1% SDS) for 1 hr. at 52°C, washed with high stringency buffer (0.1 x SSC, 1% SDS) for 1 hr. at 50°C, dried, and then exposed to X-ray film for 3 hours.

The detection of non-radioactive probes was accomplished as described above. Both the biotin-11-dUTP and the [ $\alpha^{33}\text{P}$ ]dCTP 2-4 TCL probes recognized a large molecular weight band at about 20 kb in all the

- 40 -

lanes containing digested *T. flavus* genomic DNA, although the mobility of the bands varied somewhat in the lanes containing different digests. The probes did not bind to the control IL-3A DNA, suggesting that the probes were specific for the target and could be used to screen the *T. flavus* genomic library.

To screen the amplified *T. flavus* genomic library (Example 2), the phage library was plated on two plates each at  $10^5$  plaque-forming units (pfu)/100 mm 2XTY plates. Duplicate plaque lifts on Hybond N from each plate were obtained and prepared for hybridization by methods well known in the art (Sambrook, Fritsch, and Maniatis, Molecular Cloning, A Laboratory Manual, 2nd ed., Cold Spring Harbor Press (1989)). The DNA on the plaque lifts was UV-cross-linked to the Hybond N for 3 minutes and each plaque lift was placed in a heat-sealed plastic bag and prehybridized as described above. Seven  $\mu$ l of either the 2-4 biotin-11-dUTP TCL probe or the 2-4 fluorescein-12-dUTP probe were added to one set of the plaque lifts and  $1.5 \times 10^6$  cpm of the [ $\alpha^{33}$ P]dCTP 2-4 TCL probe were added to the duplicate filters. The filters were incubated overnight at 52°C and washed the next day with low and high stringency buffers as described above. The filters with non-radioactive probes were incubated with 10 ml of conjugation buffer (0.5% casein, 1 x PBS and 0.02% Na-azide) for 30 min. at room temperature. Hybridization conditions, washes and detection were as described above.

Approximately 25 positive plaques (hybridizing with the labeled probes) out of  $10^5$  pfu from the amplified CIP Tfl DNA library were detected on the duplicate plaque lifts.

Ten positive plaques were selected ( $\lambda$ 21,  $\lambda$ 31,  $\lambda$ 51,  $\lambda$ 61,  $\lambda$ 71,  $\lambda$ 81,  $\lambda$ 91,  $\lambda$ 101,  $\lambda$ 111 and  $\lambda$ 121) and were purified by two rounds of dilution and screening with the labeled 2-4 probes, until well-isolated, single, positive plaques were obtained.

The four stocks of phage  $\lambda$ 21,  $\lambda$ 51,  $\lambda$ 71, and  $\lambda$ 91 were grown at  $5 \times 10^5$  pfu/2XTY plate, 5 plates per stock. The phages were eluted from

- 41 -

the plates by a standard protocol (Sambrook et al. (1989)). The eluant was treated with 20 $\mu$ g/ml DNase and 50 $\mu$ g/ml RNase A for 1 hr. at 37°C and extracted with both phenol-chloroform and chloroform. The DNA was ethanol-precipitated, pelleted, rinsed with ethanol, resuspended in 1 ml of TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and purified using the Lambda prep kit from CHIMERx.

The phage DNA was restriction-digested with EcoRI and BamHI and electrophoresed on a 0.7% agarose gel, transferred to Hybond N and probed with the 2-4 TCL probes. Based on agarose gel band distribution and Southern blot detection by the 2-4 probes, the four phages were grouped into two classes. Clones  $\lambda$ 21 and  $\lambda$ 91 belong to one class and the  $\lambda$ 51 and  $\lambda$ 71 belong to a second class. Clones  $\lambda$ 21 and  $\lambda$ 51 were chosen for further analyses.

Clone  $\lambda$ 21 was digested with BamHI and the *T. flavus* insert was subcloned into pTZ18U (Mead et al., *Protein Engineering* 1: 67-74 (1986)). Eight of these clones were sequenced using the SEQUAL™ Sequencing kit from CHIMERx. One of these clones, designated p21BG, hybridized to the 2-4 TCL probes and yielded sequences identical to the sequence of the 2-4 fragment between the BamHI and Eco47III sites (these sites begin at positions 2084 and 2387 in Fig. 4, respectively). This sequence information confirmed that the clones contained authentic *T. flavus* DNA polymerase gene sequence, and confirmed the orientation of this gene sequence in the clones.

Based on agarose gel analysis neither  $\lambda$ 21 nor  $\lambda$ 51 had any internal Eco RI sites, hence  $\lambda$ 21 and  $\lambda$ 51 were restriction-digested with Eco RI and the insert was cloned into pTZ18U for ease in further analyses. The resulting recombinants were designated p21E10 and p51E9, respectively (FIGURE 1A). Each clone had an insert of about 14-16 kb.

## EXAMPLE 5

Sequencing the *T. flavus* DNA Polymerase I Gene

Three main strategies were adopted for sequencing the Tfl DNA pol I gene.

5 In a first strategy, several primers were designed based on the purported *T. flavus* DNA pol I sequence published by Akhmetzjanov and Vakhitov (1992) and were synthesized by Synthetic Genetics, San Diego, CA (e.g., primers FTFL10, FTFL11, RTFL12, RTFL13, FTFL15, RTFL16, FTFL17 and RTFL18 (Table 2)). Dideoxy- sequencing of the  $\lambda$ 21 and  $\lambda$ 51  
10 clones was attempted using these primers, but only primers FTFL17 and RTFL18 yielded good sequence data and only very faint bands were obtained with primers FTFL11 and RTFL13, suggesting only partial homology to this purported DNA pol I sequence.

In a second sequencing strategy, deletion vectors of p21E10  
15 were obtained by restriction digestion of the plasmid with HincII, HindIII, SphI, KpnI and BamHI. These restriction enzymes cut once in the multiple cloning site and once or twice in the insert. The digests were diluted to allow self-ligation and transformed into *E. coli* strain DH5 $\alpha$ F' by standard methods. The clones that ligated back to the vector were selected on ampicillin-  
20 containing plates and picked for further sequence analysis.

All of these clones had deletions of different lengths at the 3' end. The size of the insert in the HincII deletion vector (p21EHc) was approximately 4.6 - 4.7 kb; in the KpnI deletion vector (p21EK) about 7 kb; in the HindIII deletion vector (p21EHd) about 1.4 kb; in the SphI deletion  
25 vector (p21ES) about 1.6 kb; and in the BamHI deletion vector (p21EB) about 1.2 kb. The plasmids p21EHd, p21EB and p21ES were sequenced with [ $\gamma$ -<sup>32</sup>P] end-labeled FSP by dideoxy sequencing. The sequences obtained were within the region of the 2-4 fragment, further confirming the orientation of the



- 43 -

insert and the presence of the 3' region of the *T. flavus* DNA polymerase I in the p21E10 parent clone.

Clone p21EHc is a deletion derivative containing the entire portion of the Tfl DNA polymerase I gene DNA present in p21E10 and about 3 kb DNA downstream from the stop codon of the Tfl DNA pol I gene, but lacking about 9.0 kb of unwanted 3' end sequence (FIGURE 1A). In addition, the DNA sequence obtained from p21E10 and p21EHc using [ $\gamma$ -<sup>32</sup>P] end-labeled reverse sequencing primer (RSP, Table 2) suggested that p21E10 contained only about 2/3 of the DNA polymerase gene and lacked the 5' one-third of the gene. In contrast, sequence obtained from p51E9 suggested that this clone contains a 5' portion of the Tfl DNA pol I gene that overlaps the coding sequence contained in p21E10, as well as significant Tfl DNA upstream of the gene.

A primer walking sequencing strategy was employed to obtain the remainder of the sequence of the Tfl DNA pol I gene. This strategy is described as "Directed Sequencing with Progressive Oligonucleotides" in Sambrook et al., Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Press (1989). To obtain additional sequence information from the clones by primer walking, primers homologous or complimentary to the ends of previously determined sequences obtained from p21E10, from the deletion vectors, and from primers FTFL17 and RTFL18 were synthesized as described above and used in additional sequencing reactions. By repeating this process the entire length of the gene was sequentially sequenced.

Specifically, TFL primers FTFL17A, RTFL18A, TFLEF1 and TFLER1 (Table 2 and FIGURE 4) were synthesized by Synthetic Genetics for primer walking based on the sequence information obtained from primers FTFL17, RTFL18 and RSP on the p21E10 template DNA (Table 2 and FIGURE 4). Primer TFLSF1 was designed as a forward primer for walking into the 3' end of the DNA pol I gene, by utilizing sequence information from p21ES. As more information became available, additional TFL primers

RTFLA - FTFLZ (Table 2; FIGURE 4) were designed for sequencing both strands.

The DNA sequence of the *T. flavus* DNA polymerase I gene and flanking sequences are given in FIGURE 2 (SEQ. ID NO. 1), along with the deduced amino acid sequence (SEQ. ID NO. 2). The sequence of 3326 b.p. has been determined. FIGURE 2 depicts 3048 bases of this sequence, of which 2505 bases are deduced to encode a polypeptide of 834 amino acids (plus stop codon). The coding sequence was determined to be 86% homologous to the Taq polymerase gene and 83% homologous to the purported Tfl polymerase gene published by Akhmetzjanov and Vakhitov.

The deduced amino acid sequence of the *T. flavus* DNA polymerase I gene was aligned and compared to the deduced amino acid sequences of the purported Tfl DNA pol I published by Akhmetzjanov and Vakhitov (85% homology) and to the deduced amino acid sequence of the Taq pol I gene (87% homology). As shown in FIGURE 3, the amino acid alignment chosen to maximize homology reveals two single amino acid insertions in the *T. flavus* DNA pol I reported here, relative to the other two reported sequences. The single amino acid inserts are depicted by dashes (-) in the sequences for Taq pol I and for Akhmetzjanov and Vakhitov's purported Tfl DNA pol I.

- 45 -

TABLE 2

<u>Primer name:</u>		<u>Primer Sequence</u>	<u>Seq. Id.</u>
		<u>No:</u>	
5	FSP	CGC CAG GGT TTT CCC AGT CAC GAC	6
	RSP	AGC GGA TAA CAA TTT CAC ACA GGA	7
	FTFL2	CTA AGT AGC TCC GAT CCC AAC	8
	RTFL4	ATC ACT CCT TGG CGG AGA GCC AGT C	9
	FTFL10	ATT TAG CAC ATA TGG CGA TGC TTC CC	10
	FTFL11	CTT TCC AGC TCC GAC CCC AAC	11
10	RTFL12	CCT ACT CCT TGG CGG AGA GCC AGT C	12
	RTFL13	TGG ATG TCC CTC CCC TCA AAG A	13
	FTFL15	CGC TTT CCC GGA AGC TTT CCC AGG TGC A	14
	RTFL16	TGC ACC TGG GAA AGC TTC CGG GAA AGG G	15
	FTFL17	CCT GCA GTA CCG GGA GCT CAC CAA GCT CAA	16
15	RTFL18	TTG AGC TTG GTG AGC TCC CGG TAC TGC AGG	17
	FTFL17A	TGG ACT ATA GCC AGA TAG AGC T	18
	RTFL18A	AAG CGA AGA CCT CCT CCT CGA	19
	TFLEF1	AGT TCG GCA GCC TCC TCC ACG A	20
	TFLEF1	TCC AAG GAA AGC CTG AGG TCT T	21
20	TFLSF1	AAG CTC GCC ATG GTG AAG CTC TT	22
	RTFLA	TCG GAG ACG AGT TGG TAG AGG T	23
	FTFLB	ACC TCT ACC AAC TCG TCT CCG A	24
	RTFLC	AGA GGA CGA AGC CCA CGA A	25
	RTFLD	AGG AGG TAG GCG AGG AGC AT	26
25	FTFLE	ATG CTC CTC GCC TAC CTC CT	27
	FTFLF	TCG AGG AGG AGG TCT TCG CTT	28
	RTFLG	AGC TCT ATC TGG CTA TAG TCC A	29
	FTFLH	ATA GGC TCT CCC AGG AGC TT	30
30	RTFLI	AAG AGC TTC ACC ATG GCG AGC TT	31
	RTFLJ	TTT CCC TGG AGG CGT TTC TGA	32
	RTFLK	AAA GAC CAC GAA GAC GCC CTT	33
	FTFL	AAG GCC GTC TTC GTG GTC TTT	34
	FTFLM	AAG GAG TGG GGA AGC CTG GAA	35
	RTFLN	TTT CAG GCT TCC CCA CTC CTT	36
35	RTFLO	TTT TTC CGA AGA GGG TTT CCA	37
	RTFLP	CGG TCC AGG AGC GCC CTG AGG A	38
	FTFLQ	CCT CAG GGC GCT CCT GGA CGC CA	39
	FTFLR	TTT GTC CTC TCC CGC CCC GA	40
	FTFLS	CCA ACC TGC AGA ACA TCC CCG T	41
40	RTFLT	GGT GTG GAT GTC CTT CCC CT	42
	FTFLU	CCC TGC CGT TTA GAG GAA GTT CAA G	43
	RTFLV	CTT GAA CTT CCT CTA AAC GGC AGG G	44
	RTFLW	ACC CGG CCT TTG GGT TCA AAG A	45
	FTFLX	TCT TTG AAC CCA AAG GCC GGG T	46
45	RTFLY	TTT CCG TGC TCC TTC CGC TC	47
	FTFLZ	CTC GCC TTC CTC GTG CCC TT	48
	5' lac PCR	GCT TCC GGC TCG TAT GTT GTG TG	49
	TFL-SDM-1	GGA AAG CCT GAG GTC TTC CAT AGC TGT TTC CTG 50	
		TGT GAA ATT GTT ATC CGC TCA CAA TTC CAC ACA	
50	TFL-SDM-3	ACA T	
		ACC CGG CCT TTG GGT TCA AAG AGC GGA ACG ATC 51	
		GCC TCC ATA GCT GTT TCC TGT GTG AAA TTG TTA	
		TCC GCT CAC AAT TCC	

## EXAMPLE 6

### Construction and Expression of a Full-Length *T. flavus* DNA Pol I Clone and Purification of Full-length Recombinant *T. flavus* DNA Pol I

5           An expression vector containing the full-length *T. flavus* DNA polymerase I gene was constructed as described below, utilizing plasmid p51E9, which contains the 5' portion of the gene, and plasmid p21E10, which contains a 3' portion of the gene that overlaps the 5' portion contained in p51E9. FIGURES 1A and 1B are provided to illustrate steps in the  
10       construction of expression vectors of this invention, and are not intended to be a scale representation of clone inserts, or to contain a complete restriction map of clones depicted therein for enzymes shown.

          Referring to FIGURE 1A, clone p51E9 which carries the 5' portion of the Tfl DNA pol I gene, was digested with BamHI and a 3.7 kb  
15       digestion product was subcloned into the BamHI cloning site of pTZ18U to produce recombinant plasmid, p51B4, which was characterized as containing about 1.5 kb of DNA upstream of the DNA pol I start codon contiguous with the 5' region of the Tfl DNA pol I gene extending to the BamHI site in the 2-4 fragment. Plasmid p51B4 was then digested with XbaI, and a 2.5 kb  
20       digestion product was subcloned into the XbaI site of pTZ18U to create plasmid p51X16, which contained only approximately 0.3 kb of DNA upstream of the DNA pol I start codon.

          Next, plasmid p21EHc (a subclone of p21E10 described above) was digested with BamHI and SalI. The 3.7 kb fragment containing the 3'  
25       region of the Tfl DNA pol I gene, beginning with the BamHI site in the 2-4 fragment, was isolated and subcloned into pTZ18U that had been digested with BamHI and SalI to create clone p21BHc. Clone p21BHc was digested with EcoRV and BamHI and the 1.3 kb fragment containing the 3' region of Tfl DNA pol I was ligated into pTZ18U that had been digested with BamHI  
30       and HincII, yielding p21BRV2.

- 47 -

Plasmid p51X16 was digested with BamHI and the 2.5 kb BamHI insert was isolated. Plasmid p21BRV2 was linearized with BamHI and ligated to the BamHI fragment of p51X16. The resulting clones were designated pTFL 1.3 and pTFL 1.4. The integrity of the Tfl DNA pol I gene in clone pTFL 1.4 was verified by DNA sequence analysis using the primer RTFLG (Table 2 and FIGURE 4).

Competent *E. coli* DH5 $\alpha$ F' were transformed with plasmid pTFL 1.4 (the 1st generation expression clone), from which a Tfl DNA pol I protein was isolated and purified as follows. *E. coli* DH5 $\alpha$ F'[pTFL-1.4] were grown in a 50 liter fermentor (10 pounds back pressure, 30 lpm aeration, 200 rpm agitation, at 37°C) in TB medium (Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd ed. (1989): 12 g Bactotryptone, 24 g yeast extract, 4 ml glycerol, 0.1 g MgSO<sub>4</sub>, 2.31 g KH<sub>2</sub>PO<sub>4</sub>, 12.54 g K<sub>2</sub>HPO<sub>4</sub> per liter), supplemented with 50  $\mu$ g/ml ampicillin with vigorous aeration at 37°C. At O.D.<sub>600</sub> = 1.0, IPTG was added to a final concentration of 0.5 mM and the cells were cultured for an additional 2 hrs. The culture was cooled down to 20°C and 100 ml of 100 mM phenylmethyl sulfonyl fluoride (PMSF) in isopropanol was added. After brief mixing, the culture was spun down in a Sharples centrifuge and the pellet (or paste) was stored frozen at -70°C. Fifty grams of cells were thawed in 250 ml of lysis buffer A (20 mM Tris-HCl pH 7.4, 0.5 mM EDTA, 100 mM NaCl, 5% glycerol, 5 mM  $\beta$ -mercaptoethanol, 0.5% Nonidet P40, 0.5% Tween 20, 50  $\mu$ g/ml PMSF, 0.5  $\mu$ g/ml pepstatin A, 0.5  $\mu$ g/ml leupeptin). The cell suspension was homogenized twice in a Manton-Gaulin press. Because PMSF is unstable in aqueous solutions, new PMSF was added again to a final concentration of 50  $\mu$ g/ml after the first and second homogenizations.

The suspension of broken cells was divided into 100 ml aliquots and heated to 65°C for 1 hr. to denature the bulk of *E. coli* proteins, including nucleases, proteases and *E. coli* polymerases. Cell debris and denatured proteins were centrifuged at 6,800 x g for 30 min. and the NaCl

concentration of the supernatant was adjusted to 400 mM. The presence of DNA polymerase activity in the supernatant was confirmed using the standard activity assay described above. Then 10% polyethyleneimine (PEI, pH 7.5) was slowly added to a final concentration of 0.2%. After 30 min. of stirring at 4°C, the suspension was centrifuged (1 hr., 6,800 x g) and the resulting supernatant was diluted with 6 volumes of buffer C (20 mM Tris-HCl, pH 7.4, 0.5 mM EDTA, 5% glycerol, 5 mM  $\beta$ -mercaptoethanol, 0.5% Nonidet P40, 0.5% Tween 20, 50  $\mu$ g/ml PMSF, 0.5  $\mu$ g/ml pepstatin A, 0.5  $\mu$ g/ml leupeptin). Ammonium sulphate was added to 0.55 g/ml and the mixture was stirred slowly overnight at 4°C.

After centrifugation for 2 hrs. at 6,800 x g the supernatant was carefully removed and tested for DNA polymerase activity. The polymerase-containing pellet was dissolved in a total of 80 ml of Buffer A (10 mM KPO<sub>4</sub>, pH 7.0, 0.5 mM EDTA, 100 mM NaCl, 5 % glycerol, 5 mM  $\beta$ -mercaptoethanol, 0.5% Nonidet P40, 0.5% Tween 20, 50  $\mu$ g/ml PMSF). Insoluble material was removed by centrifugation at 6,800 x g for 20 min. The supernatant obtained from this centrifugation (which contains the polymerase activity) was loaded onto a 5 x 50 cm Sephadex G-25 column equilibrated in buffer A to desalt the solution and to remove traces of PEI. The flow rate used on this column was about 200 ml/hr. Fractions of 25 ml were collected and assayed for DNA polymerase activity. The flow-through fractions contained the activity. It was essential to remove all the PEI for efficient adsorption to the next column.

The crude Tfl DNA polymerase described above was applied to a Bio-Rex 70 column (5 x 10 cm) (Bio-Rad) equilibrated in Buffer A. The column was washed with 1.5 l of buffer A and the DNA polymerase was eluted with 4 l of a 0 - 1 M NaCl gradient in buffer A. Fractions of 25 ml were collected and assayed for DNA polymerase activity. Fractions containing DNA polymerase activity (as assayed below) were pooled, concentrated in an Amicon concentrator with a YM30 membrane to about 40

ml and dialyzed against two changes (1 liter each) of antibody column high salt buffer B (20 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 0.5 M NaCl, 0.05% Brij-35) and applied to an immunoaffinity column (1.5 x 8 cm).

The immunoaffinity column was prepared using techniques well-known in the art. First, a mouse is injected with purified DNA polymerase I to provide an immune response; preferred DNA polymerases for generating antibodies are thermophilic DNA polymerases, including those isolated from *Thermus flavus* and *Thermus aquaticus*. A ten week old female BALB/c mouse (Harlan Sprague Dawley, Madison, WI) was immunized by intraperitoneal injection with Taq polymerase (purified from *Thermus aquaticus* ATCC #25104). To prepare the Taq polymerase for injection, Taq storage buffer was removed with a Centricon 30 protein concentrator (Amicon Corp.), and the concentrated protein was diluted with phosphate-buffered saline. For the initial immunization, 40 µg of Taq emulsified with complete Freund's adjuvant was injected. Five booster injections of 40µg Taq polymerase mixed with equal volumes of the Ribi Adjuvant System (Ribi Immunochem Research, Inc., Hamilton, MT) were administered over a 6 month period, with successive intervals between injections of approximately five weeks, 4 weeks, 4 weeks, 12 weeks, and 4 weeks.

Five days after the final booster injection, the mouse was sacrificed and spleen cells were isolated and fused with myeloma cells (myeloma P3X63-AG8.653 (ATCC CRL 1580)) to generate hybridomas, using techniques well-known in the art. See E. Harlow and D. Lane *Antibodies, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor (1988). In particular, fusions were performed in 50% polyethylene glycol and selected in HAT medium. All hybridomas were screened as described below.

The fifth fusion experiment yielded a useful hybridoma, as selected in the following manner. The hybridomas were distributed into 96-well plates. Of 1176 wells filled, approximately 913 showed growth. To

- 50 -

initially screen these wells, an ELISA assay was employed. First, polystyrene ELISA plates were coated with *Thermus flavus* (ATCC #33923) DNA polymerase (1  $\mu$ g/ml Tfl DNA pol I (MBR lot 20229) in 100 mM Tris-HCl, pH 8.5/0.05% NaN<sub>3</sub>). Five microliter samples of supernatant from each culture were diluted into 95  $\mu$ l of Tris-buffered saline, pH 8.5/0.05 % Triton X-100 (TBST) and then incubated in the coated ELISA plates for 2 hours at room temperature. The plates were then washed with TBST. To detect positive anti-Tfl DNA polymerase cross-reactivity, a commercially available, peroxidase-conjugated goat anti-mouse IgG,  $\gamma$  chain specific (Jackson ImmunoResearch, West Grove, PA), was diluted 5000-fold in TBST, added to the ELISA plates, and incubated for 1 hour at room temperature. Positive cross-reactivity was detected colorimetrically with 3-methyl-2-benzothiazolinonehydrazone/3-dimethylaminobenzoic acid/hydrogen peroxide.

Supernatants from wells that tested even weakly positive by ELISA were further screened by immunoprecipitation of both Tfl and Taq DNA polymerases using techniques well known in the art. See Harlow and Lane, *supra*. The immunoprecipitation assay employed relies on the presence of protein A (which binds IgG) on the surface of *Staphylococcus aureus* (SAC, Sigma Chemical Co., St. Louis, MO). Since protein A does not bind strongly to a common subclass of mouse IgG, IgG<sub>1</sub>, but does bind rabbit IgG strongly, a pellet of centrifuged SAC cells was first treated with rabbit anti-mouse IgG antibodies. The pellet from 10  $\mu$ l of a 10% suspension of these cells was then incubated with 20  $\mu$ l of hybridoma culture supernatant for one hour at room temperature. The resultant SAC cells were centrifuged, washed, and resuspended in diluted Taq or Tfl polymerase. The polymerase enzyme-cell suspensions were incubated overnight at 4°C and centrifuged. The resultant supernatant was removed and tested for depletion of polymerase activity using a standard radiochemical assay.

One hybridoma, designated hybridoma 7B12, tested strongly in the ELISA assay and immunoprecipitated both Taq and Tfl DNA polymerases



(70-99% depletion in polymerase activity). More particularly, in a series of immunoprecipitations in which the polymerase concentration (Taq or Tfl polymerase) was varied, the results shown in Table 2B were obtained.

TABLE 2B  
IMMUNOPRECIPITATION RESULTS

Trial	Polymerase	Source of Monoclonal antibody	Depletion of polymerase activity
2*	0.04 units Tfl polymerase	hybridoma 7B12 supernatant, 20 $\mu$ l	> 99%
3	0.09 units Tfl polymerase	hybridoma 7B12 supernatant, 20 $\mu$ l	80 %
3	0.27 units Taq polymerase	hybridoma 7B12 supernatant, 20 $\mu$ l	79 %
4	0.30 units Tfl polymerase	1.375 $\mu$ g purified IgG from hybridoma 7B12	99 %

\*Trial 1 was unsuccessful due to an excess of polymerase enzyme relative to the amount of antibody.

In control immunoprecipitation experiments in which six anti-*E. coli* DNA polymerase I monoclonal antibodies were employed (25  $\mu$ l of hybridoma culture supernatant), 91 to 112% of the original Tfl polymerase activity was still detectable in solution (i.e., at most a 9% depletion of Tfl polymerase activity).

The monoclonal antibody from hybridoma 7B12 was further characterized and found to neutralize Taq and Tfl polymerase activity at lower temperatures (41°C). This activity assay was performed at 41°C rather than at higher temperatures (70°C) where the enzymes are more active, because the antibody itself denatures at higher temperatures.

Cells from hybridoma 7B12 were cloned three times by limiting dilution until all wells with growth tested positive in the ELISA assay (66/66 wells). The monoclonal antibody of hybridoma 7B12, a mouse IgG ( $\gamma$ 1,  $\kappa$ )

- 52 -

antibody, is commercially available from Molecular Biology Resources, Inc., Milwaukee, Wisconsin as Cat. No. 4100-01.

Subsequent experiments produced two additional anti-Taq/anti-Tfl monoclonal antibodies that may be (but have not been) used to affinity-purify DNA polymerase enzymes. In particular, two hybridomas producing anti-Tfl DNA polymerase monoclonal antibodies, formed using spleen cells from a mouse immunized with Tfl DNA polymerase, were identified using the screening procedures outlined above. In the immunoprecipitation assay, 25  $\mu$ l of supernatant from these two hybridoma cultures, designated hybridomas 10F10 and 11G4, depleted 92% and 95%, respectively, of the DNA polymerase activity from a solution containing 0.30 units of Tfl DNA polymerase.

The monoclonal antibodies from hybridoma 7B12 were coupled with Emphaze™ resin (3M, St. Paul, MN) as follows. Twenty-five milliliters of antibody solution (2 mg/ml in 0.6 M sodium citrate, 0.05 M sodium chloride, 0.05 M HEPES pH 8.6) was added to 1.25 g of Emphaze™ resin and allowed to react for 2 hrs at room temperature. Ethanolamine (1 ml of a 3 M solution, pH 9.0) was then added to quench unreacted azlactone functional groups and incubated for 1 hr. at room temperature, then overnight at 4°C. The resin was washed with and stored in PBS with 0.05% sodium azide.

The immunoaffinity column for the DNA polymerase purification was prepared with about 10 ml dead volume of the resin and washed with 300 ml of antibody column high salt buffer B (20mM Tris-HCl pH 7.5, 0.5 mM EDTA, 0.5 M NaCl, 0.05% Brij - 35). The Tfl DNA polymerase enzyme was eluted with 10 mM triethylamine (pH 11.6). Fractions (5 ml each) were collected into tubes with 0.01 volumes of 1 M HEPES. Those fractions containing the DNA pol I enzyme were identified by activity assay, pooled, and dialyzed against storage buffer S (50% glycerol, 50 mM Tris-HCl, pH 7.5 at 23°C, 5 mM DTT, 0.1 mM EDTA, 0.5% Tween 20, 0.5% Nonidet P40). The final product was stored at -20°C.

- 53 -

The above purification procedure yielded about 60,000 units of purified *T. flavus* DNA polymerase I from 50 g of *E. coli* [pTFL-1.4] cells, which is equivalent to 1,200 units/g of cells.

The protein concentration was determined by the method of Lowry using a modification of the Sigma (St. Louis, MO) Protein Assay Kit (Cat. No. P5656) with Bovine Serum Albumin as a standard. Both standard and sample were precipitated with TCA prior to the protein analysis. Using the standard activity assay, the DNA polymerase specific activity was calculated to be 79,500 U/mg protein for the recombinant Tfl holoenzyme purified as described.

#### EXAMPLE 7

##### Construction and Expression of a High-Yield, Full-Length *T. flavus* DNA Pol I Clone and Purification of Recombinant *T. flavus* DNA Pol I Holoenzyme

To increase expression of the DNA Tfl pol I gene and to increase the yield and DNA polymerase specific activity of recombinant Tfl DNA pol I, the lacZ promoter was fused directly to the ATG start codon of the Tfl DNA pol I gene using site-directed mutagenesis, the resultant improved expression plasmid was expressed, and the recombinant DNA pol I was purified using a modified procedure.

Site-directed mutagenesis of single-stranded uracil-(U-) containing DNA from p51X16 was performed using the oligonucleotide TFL-SDM-3 (Table 2). Single-stranded U-containing DNA was prepared according to the protocol provided by Bio-Rad (Hercules, CA) in their Mutagenesis Kit. The new clone, p51X16M1, had the lacZ promoter fused to about 2 kb of the 5' portion of the Tfl DNA pol I gene (FIGURE 1A). Plasmid p51X16M1 was digested with BamHI and HincII and ligated to the 1.3 kb BamHI/EcoRV fragment isolated from p21BHc, which provided the 3' region of the Tfl DNA pol I gene. The resulting plasmid, pTFLRT4 (ATCC

Accession No. 69633), was used to transform *E. coli* DH5 $\alpha$ F'IQ (Life Technologies, Grand Island, NY), generating the 2nd generation expressing clone (FIGURE 1A). The presence and integrity of the Tn DNA pol I gene in the insert of pTFLRT4 was confirmed by DNA sequence analysis using  
5 primers RTFLJ, RTFLG, FTFLB, and FTFLF as set out in Table 2.

*E. coli* DH5 $\alpha$ F'IQ transformed with pTFLRT4 were grown in a 250 liter fermentor in TB medium supplemented with 50  $\mu$ g/ml ampicillin. At O.D.<sub>600</sub> = 0.7, expression of the plasmid was induced by the addition of IPTG to a final concentration of 0.5mM and the cells were cooled down to  
10 20°C, harvested three hours later, and stored at -70°C until use.

Five hundred grams of induced cells were thawed and suspended in 2500 ml of lysis buffer A (20 mM Tris-HCl, pH 7.4, 0.5 mM EDTA, 100 mM NaCl, 5% glycerol, 5 mM  $\beta$ -mercaptoethanol, 0.5% Nonidet P40, 0.5% Tween 20, 50  $\mu$ g/ml PMSF, 0.5  $\mu$ g/ml pepstatin A, 0.5  $\mu$ g/ml  
15 leupeptin).

The cell suspension, to which lysozyme was added to a concentration of 0.5 mg/ml, was homogenized in a Manton-Gaulin press. Fresh PMSF again was added to the lysed cells to a final concentration of 50  $\mu$ g/ml. The suspension of lysed cells was divided into 300 ml portions,  
20 heated to 65°C for 1.5 hrs., and centrifuged for 30 min. at 6,800 x g.

Following this centrifugation, the resulting supernatant was adjusted to an additional NaCl concentration of 400 mM and 10% PEI, pH 7.5, was added to a final concentration of 0.2%. After 1 hr. of stirring at 4°C, the suspension was centrifuged (1 hr., 6,800 x g) and the resultant  
25 supernatant was precipitated with ammonium sulphate. After centrifugation for 2 hours at 6,800 x g the resultant pellet was resuspended with 200 ml Buffer A and applied to a Bio-Rex 70 column (5 x 10 cm) (Bio-Rad). The column was washed with 1.5 l of buffer A and the DNA polymerase protein was eluted with 4 l of a 0-1 M NaCl gradient in buffer A. Fractions of 25 ml  
30 were collected, and the peak fractions were pooled and dialyzed against two

- 55 -

- changes (2.5 l each) of antibody column high salt buffer B and applied to an immunoaffinity column (1.5 x 8 cm) prepared as described above. After washing the immunoaffinity column with 250 ml antibody column high salt buffer B, the enzyme was eluted with 10 mM triethylamine (pH 11.6).
- 5 Fractions of 5 ml were collected and the peak fractions were dialyzed against storage buffer S. This procedure yielded about 2,000,000 units of purified *T. flavus* DNA polymerase from 500 g of *E. coli* [pTFLRT4] cells, or about 4,000 units/g of cells as measured using the standard assay described above. The calculated DNA polymerase specific activity was 217,600 U/mg for this
- 10 preparation of Tfl holoenzyme.

- The N-terminal amino acid sequence from recombinant DNA pol I holoenzyme isolated from *E. coli* [pTFLRT4] was determined as Met-Glu-Ala-Ile-Val-Pro-Leu-Phe-Glu-Pro. This sequence matches the amino terminal sequence deduced by translation of the *T. flavus* DNA pol I gene
- 15 sequence (FIGURE 2), indicating that the translation starts at the predicted position. Unlike the native holoenzyme studied, no blockage of the terminal methionine in the cloned holoenzyme was observed.

### EXAMPLE 8

#### 20 Cloning and Expression of the Exo<sup>-</sup> Fragment of *T. flavus* DNA Polymerase I

- Expression studies using plasmids p21E10 and p21EHc were performed because these plasmids contain the 3' two-thirds of the DNA polymerase I gene fused to the lacZ operator/promoter. As deduced from the DNA sequence, the first amino acid encoded by the insert of plasmid p21E10
- 25 corresponds to Glu<sub>239</sub> in FIGURE 2 (circled). It was hypothesized that the insert in p21E10 would encode a fragment of DNA polymerase I lacking the exonuclease domain (the exo<sup>-</sup> fragment) due to the absence of the 5' one-third portion of this gene.

- 56 -

From translation of sequence information obtained from the 5'-end of p21E10 using primer RSP (Table 2), it was concluded that the insert encoding the 3' two-thirds of the Tfl DNA pol I gene was out-of-frame. It was assumed that the same out-of-frame fusion was present in p21EHc. However, in spite of frame shift some heat-stable DNA polymerase activity was obtained from the clone harboring p21E10.

The ATG start codon of lacZ was brought in frame with the TFL DNA polymerase exo' fragment in p21EHc using site-directed oligonucleotide mutagenesis (FIGURE 1B). A mutagenic oligonucleotide TFL-SDM-1 was designed (Table 2), part of TFL-SDM-1 having homology to nucleotides 1015-1032 in FIGURE 4, the other part having homology to the vector. Single-stranded U-containing DNA was prepared by standard procedures and the chemically synthesized oligonucleotide TFL-SDM-1 was used to obtain site-directed changes in the newly synthesized DNA. This DNA was used to transform competent *E. coli* DH5 $\alpha$ F'. Several transformants were selected and grown up for plasmid analysis. Of these, sequence analysis was performed on four clones using the [ $\gamma^{32}$ P] end-labeled primer "5' lac PCR" (Table 2) (Synthetic Genetics). The clones with the DNA polymerase gene fragment in the proper reading frame (which includes the ATG from the lacZ coding sequence, followed by "GAA GAC..." derived from the Tfl DNA pol I gene - see FIGURE 2, nucleotides 1015-1020 et seq.) were included in expression studies. Overexpressed recombinant protein was then isolated and purified from *E. coli* transformed with one of the clones, p21EHcM1.1 (ATCC Accession No. 69632) by following the procedures outlined below.

*E. coli* DH5 $\alpha$ F' [p21EHcM1.1] was grown in a 50 liter fermentor in TB medium (Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd ed. (1989)) supplemented with 50  $\mu$ g/ml ampicillin with vigorous aeration at 37°C. At O.D.<sub>600</sub> = 1.0, IPTG was added to final 0.5 mM concentration and cells were cultured for an additional 2 hours. The culture

- 57 -

was cooled down to 20°C and 100 ml of 100 mM PMSF in isopropanol was added. After brief mixing, the culture was spun down in a Sharples centrifuge and stored frozen at -70°C.

- 5 of lysis buffer A (20 mM Tris-HCl pH 7.4, 0.5 mM EDTA, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 5% glycerol, 5 mM β-mercaptoethanol, 0.5% Nonidet P40, 0.5% Tween 20, 50 μg/ml PMSF, 0.5 μg/ml pepstatin A, 0.5 μg/ml leupeptin). The cell suspension was homogenized twice in a Manton-Gaulin press. After the first and second passes, fresh PMSF was added again to a
- 10 new, final concentration of 50 μg/ml. The suspension of broken cells was divided into 100 ml portions and heated to 65°C for 1 hr. Cell debris and denatured proteins were centrifuged at 6,800 x g for 30 min. and the supernatant was adjusted to an additional NaCl concentration of 400 mM. Then 10% PEI, pH 7.5, was slowly added to a final concentration of 0.2%.
- 15 After 30 min. of stirring at 4°C, the suspension was centrifuged (1 hr., 6,800 x g) and the supernatant was concentrated on a YM30 membrane to 100-120 ml. The concentrate was run through a 5 x 50 cm Sephadex G-25 column equilibrated in buffer A, as described in Example 6. The crude Tfl exo' fragment was applied to a Procion-Red Sepharose column (5 x 10 cm). The
- 20 column was washed with 1.5 liters of buffer A and the DNA polymerase fragment was eluted with 4 liters of a 0-1.5 M NaCl gradient in buffer A. Fractions of 25 ml were collected and the fractions with DNA polymerase activity were dialyzed against two changes (1 liter each) of antibody column high salt buffer B (20 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 0.5 M NaCl,
- 25 0.05% Brij-35) and applied to an immunoaffinity column (1.5 x 8 cm). After washing the column with 250 ml of the same buffer, the enzyme was eluted with 10 mM triethylamine (pH 11.6) and treated as described above. In general, about 300,000 units of purified *T. flavus* exo' fragment were obtained from 50 g of *E. coli* [p21EHcM1.1] cells (6,000 units/g).

- 58 -

The protein concentration was determined as described above. The calculated DNA polymerase specific activity for the Tfl exo' fragment was 600,000 U/mg.

Once the Tfl exo' fragment was cloned and expressed the N-terminal amino acid sequence was determined. About 50  $\mu$ g of the purified enzyme was separated using SDS-PAGE and blotted onto PVDF membrane as described for the holoenzymes. The major band was excised and subjected to sequence analysis. The chromatogram of the sequencer indicated the presence of a major and a minor sequence. The minor sequence represents the major sequence shifted by one amino acid. The major sequence reads: Leu-Glu-Arg-Leu-Glu-Phe-Gly-Ser-Leu-Leu-His-Glu-Phe-X-Leu-Leu-X-Ala-Pro-Ala (where X represents an amino acid whose identity was uncertain from the chromatogram). The minor sequence has the amino acid sequence: Glu-Arg-Leu-Glu-Phe-Gly-Ser-Leu-X-His-Glu-Phe-Gly-X-X-Pro-X-X-Ala-Pro.

The major sequence is identical to the amino acid sequence deduced from the recombinant Tfl exo' fragment DNA sequence, except for the lack of 37 N-terminal amino acids, including the N-terminal methionine. SEQ ID NO: 3 and 4 contain the DNA sequence and the deduced amino acid sequence of the Tfl exo' fragment, as expected from construct p21EHcM1.1. The loss of the 37 N-terminal amino acids may be due to processing of the exo' fragment in the *E. coli* host. SEQ ID NO: 5 contains the amino acid sequence of the major band exo' fragment, as deduced from the N-terminal amino acid sequence of the purified exo' fragment and from the DNA sequence of plasmid p21EHcM1.1. The minor sequence presented here is the Tfl exo' fragment lacking both the N-terminal methionine and the next 37 amino acids. Although the amount sequenced in the minor species was small there was good correlation with the deduced amino acid sequence, except for the proline at position 16, that was expected to be glutamic acid.



## EXAMPLE 9

Characterization of *T. flavus*  
DNA Polymerase I Exonuclease Activities

The purity and molecular weight of the *T. flavus* DNA polymerase and the Tfl exo<sup>-</sup> fragment were estimated by SDS-polyacrylamide gel electrophoresis using the Pharmacia PhastSystem (Piscataway, NJ). FIGURE 9 shows the purity of the holoenzyme and the Tfl exo<sup>-</sup> fragment, which were separated on a 12.5% SDS-PAGE gel and stained with silver.

Assays were performed to determine intrinsic/extrinsic exonuclease, endonuclease, and DNase activities of the DNA polymerase enzyme preparations purified as described above and for *T. aquaticus* DNA pol I holoenzyme (Taq holo) and Stoffel fragment (Stoffel, Perkin Elmer, Foster City, CA, Cat. No. N808-0038), and for *T. thermophilus* holoenzyme (Tth holo) (Molecular Biology Resources, Inc., Cat. No. 1115-01, Milwaukee, WI). The protocols are described below and results summarized in Table 3A.

A 3' → 5' exonuclease activity assay was performed in a final volume of 10 μl containing 50 mM Tris-HCl, pH 7.6, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 0.15 μg of [3'-<sup>3</sup>H dCTP and dGTP labeled] λ DNA/Taq I fragments and 5, 10 and 20 units of enzyme. Each sample was overlaid with 10 μl of light mineral oil and incubated at 70°C for 1 hour. The reaction was terminated by the addition of 50 μl yeast RNA and 200 μl of 10% TCA. After incubation for 10 min. on ice, the samples were centrifuged for 7 min. in a microcentrifuge. 200 μl of supernatant was added to 6 ml of scintillation fluid and counted in a scintillation counter. The results are presented in Table 3A as the slope %-label released per unit of enzyme.

A 5' → 3' assay was performed in a manner identical to the 3' → 5' exonuclease assay, except for the use of [5'-<sup>32</sup>P] λ DNA/ HaeIII fragments instead of the 3'-labeled substrate.

- 60 -

Double-stranded and single stranded DNase assays were performed using the protocol for the 3' → 5' exonuclease assay, except for the use of [<sup>32</sup>P] λ DNA instead of the 3'-labeled substrate. The DNA was treated for 3 min. at 100°C and immediately chilled on ice prior to assaying for single stranded DNase activity.

An assay for endonuclease activity was performed as follows. The reagents (final concentrations of 50 mM Tris-HCl, pH 7.6, 10 mM MgCl<sub>2</sub>, 1 mM β-mercaptoethanol), 0.5 μg pBR322, less enzyme/H<sub>2</sub>O, were mixed and kept on ice. The required amount of H<sub>2</sub>O and 10 μl mineral oil were added to each tube. The reaction was started with the addition of 5, 10 or 20 units of enzyme; the final reaction volume was 10 μl. The samples were incubated at 70°C for 1 hour. Two μl of a solution containing 0.25% bromophenol blue, 1 mM EDTA, and 40% sucrose was added to the reaction, and after a short centrifugation, 6 μl of the bottom layer was removed and electrophoresed on 1.5% agarose gels in 1 x TBE. The mobility change from the supercoiled to the linear form of pBR322 was recorded.

TABLE 3A

Enzyme	3'→5' exo-nuclease	5'→3' exo-nuclease	ss DNase	ds DNase	Endo-nuclease
Tfl holo (r)	0.19	0	0.66	0	0
Tfl holo (n)	0.04	0.01	0	0	0
Tfl exo	0.03	0.002	0	0	0
Taq holo	0.031	0.09	0	0	0
Stoffel	0	0.01	0.28	0.1	0
Tth holo	0.07	0.04	0.2	0	0

The values for 3'→5' exonuclease activity and for 5'→3' exonuclease activity are low for all DNA polymerases tested. The differences in exonuclease and DNase activities between naturally occurring and recombinant Tfl holoenzyme are not believed to be statistically significant.

#### EXAMPLE 10

##### Comparison of the *T. flavus* and *T. aquaticus* DNA Polymerases

Biological properties of native *T. flavus* DNA pol I (nTfl Holo, lot #30419; Molecular Biology Resources, Inc., Cat. No. 1112-01, Milwaukee, WI); recombinant *T. flavus* DNA pol I holoenzyme (rTfl Holo purified from *E. coli* [pTFLRT4]; *T. flavus* DNA pol I exo' fragment (Tfl exo' ) purified from *E. coli* [p21EHcM1.1]; *T. aquaticus* DNA polymerase I (native Taq or recombinant AmpliTaq) holoenzyme; and the AmpliTaq DNA polymerase Stoffel fragment were compared using a number of protocols described below.

The molecular weights and purities of the preparations of the various enzymes were estimated by acrylamide gel electrophoresis utilizing the Pharmacia PhastSystem (Piscataway, NJ) for electrophoresis and silver staining. A comparison of the apparent molecular weights estimated from 7.5% and 12.5% acrylamide gels and the calculated molecular weights derived from available sequence data is given in Table 3B. The apparent molecular weight of the holoenzymes using either acrylamide concentration was less than the calculated molecular weights. A purity of greater than 95% was estimated for all DNA polymerases analyzed: i.e. Tfl and Taq holo enzymes, Tfl exo' fragment and Stoffel fragment.

- 62 -

**TABLE 3B**  
Apparent Mol. Weight

Enzyme	7.5% gel	12.5% gel	Calculated Mol. Weight
Tfl holoenzyme*	80,000	84,000	93,969
Recombinant Tfl exo' fragment (recomb.)	59,000	59,000	62,979
Taq holoenzyme	82,000	85,000	93,904
Stoffel	60,000	61,000	61,000

\* native and recombinant

Using the Pharmacia PhastSystem, the polymerases and standards were subjected to isoelectric focusing. The experimentally derived pI values of the samples, including samples of *E. coli* DNA pol I holoenzyme (*E. coli* pol I) and Klenow fragment, were compared to values calculated from derived amino acid sequence information. The results are given in Table 4.

**TABLE 4**  
pI Values

Enzyme	Calculated pI	Measured pI
nTfl Holo	6.23	6.25
rTfl Holo	6.23	6.43
A&V Tfl'	5.73	(not available)
Tfl exo'	6.37	5.94
Taq Holo	6.00	6.14
Taq Stoffel	5.93	5.83
<i>E. coli</i> pol I	5.29	5.12
Klenow	5.60	5.75

\*Purported *T. flavus* DNA pol I protein sequence published by Akhmetzjanov and Vakhitov.

- 63 -

The relative DNA polymerase activities of the enzymes were assayed at 70°C at different pH values. The pH of selected buffers were adjusted at 23°C, to permit direct comparison to published results. Table 5A shows the measured pH values at 70°C for 1x buffers which first had been titrated at 23°C. Unless otherwise indicated, pH values reported herein were adjusted at about 23°C.

**TABLE 5A**  
Change of pH as a function of temperature

No.	Buffer	pH at 23°C	pH at 70°C
1.	PIPES-NaOH	6.0	5.5
2.	PIPES-NaOH	6.5	6.0
3.	Tris-HCl	7.5	6.4
4.	Tris-HCl	8.0	7.0
5.	Tris-HCl	8.5	7.4
6.	Tris-HCl	9.0	8.0
7.	Tris-HCl	9.5	8.6
8.	Triethylamine-HCl	9.5	8.9
9.	Triethylamine-HCl	10.0	9.15

The activity assays were performed in a 100  $\mu$ l (final volume) reaction mixture, containing 0.1 mM dCTP, dTTP, dGTP, [ $\alpha^{32}$ P]dATP, 0.3 mg/ml activated calf thymus DNA and 0.5 mg/ml BSA in a set of buffers containing: 50 mM KCl, 1 mM DTT, 10 mM  $MgCl_2$  and 50 mM of one of three buffering compounds: PIPES, Tris or Triethylamine. Three dilutions (20, 40 or 80 U/ $\mu$ l) of each polymerase enzyme were prepared, and 5  $\mu$ l of a dilution was added to the reaction mixture, followed by incubation at 70°C for 30 min. The experiment was performed twice, each time using duplicate samples. FIGURE 5 graphically depicts the relative activities of the enzymes

- 64 -

studied, calculated as the ratio of counts per minute (corrected for background and enzyme dilution) at a given pH to counts per minute at the maximum value for that enzyme. The optimal ranges (> 80% activity) for the five enzymes tested are provided in Table 5B.

5

10

TABLE 5B Optimal pH ranges (as titered at 23°C)	
Enzyme	pH
Native Tfl holoenzyme	9.5-10.5
Recombinant Tfl holoenzyme	9.5-10.5
Tfl exo' fragment	7.5-9.8
Stoffel fragment	7.5-9.8
Ampli Taq	7.5-9.3

These values are about 1 pH unit higher than for buffers measured at 70°C (see Table 5A).

15

20

25

The pH protocol described above was modified to determine the influence of  $MgCl_2$  concentration on the activities of the DNA polymerases. The reaction buffers included 50 mM Tris-HCl pH 8.3 (23°C) and  $MgCl_2$  concentrations from 0.36 to 50 mM. Three independent experiments were performed and curves were constructed (FIGURE 6A) showing the relative activity of Tfl exo' fragment, Tfl holoenzyme (native and recombinant), and Taq Stoffel fragment. The higher limit for the Stoffel fragment was extrapolated. The optimal ranges (> 80% activity) are 1.3-13 mM  $MgCl_2$  for the Tfl exo' fragment, and 2.3-33 mM  $MgCl_2$  for the Stoffel fragment. The recombinant and the native Tfl holoenzyme showed greatest activity at 50 mM  $MgCl_2$ .

The above protocol was modified to determine the influence of  $MnCl_2$  concentration on the activities of the DNA polymerases (in the absence of magnesium ions). The reaction buffers included  $MnCl_2$  concentrations

- 65 -

from 0.1 to 20 mM. Due to the precipitation of oxidation products ( $MnO_2$ ) of  $MnCl_2$ , the reaction buffers were prepared just prior to the assay. The pH of the buffer was adjusted to pH 8.7 before the addition of a 1 M stock solution of  $MnCl_2$ . The pH was finally adjusted to 8.3 at 23°C. Three independent experiments were performed and a curve was constructed (FIGURE 6B) showing relative activity of the enzymes. The optimal ranges for the four enzymes tested are 2.1-11 mM  $MnCl_2$  for the Tfl *exo'* fragment, 4-20 mM  $MnCl_2$  for the Stoffel fragment and 0.8-4 mM  $MnCl_2$  for the recombinant and native Tfl holoenzymes.

The thermostability and temperature optimum of the polymerase enzymes were determined by incubating 10 units of enzyme for 30 min. at 23, 37, 60, 65, 70, 75, 80, and 90°C, in 100  $\mu$ l of buffer used for the determination of polymerase activity (including 50 mM Tris-HCl pH 8.3 (23°C) and 1.5 mM  $MgCl_2$ ) in a DNA polymerase activity assay as described above. The polymerase activity was then determined by acid precipitation of the polymerization product as described above. FIGURE 7A depicts the percent relative activity, calculated as described above. The temperature optima were 70-75°C for the Stoffel and Tfl *exo'* fragments and 80°C for the native and recombinant Tfl holoenzymes. At 90°C there was 14%, 6% and 8% of the activity left in the Tfl holoenzymes, the Stoffel fragment, and the Tfl *exo'* fragment, respectively.

The PCR half lives of the enzymes were determined in 100  $\mu$ l PCR reactions, performed in duplicate, substituting the appropriate buffer in the PCR cocktail prepared for each individual enzyme. The cocktail for the Tfl *exo'* fragment contains 1 x Tfl pol buffer (50 mM Tris-HCl, pH 9.0 at 23°C, 20 mM  $(NH_4)_2SO_4$ , 1.5 mM  $MgCl_2$ ), 200  $\mu$ M of each dNTP, 0.5  $\mu$ M of primer FTFL2 and primer RTFL4, and 15 ng of *T. flavus* genomic DNA. The buffers for other enzymes tested were as follows: Taq pol I (1 x Taq pol buffer: 10 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM  $MgCl_2$ ); Tfl DNA pol I holoenzyme (1 x Tfl pol buffer); and Stoffel fragment (1 x Stoffel buffer: 10

- 66 -

mM Tris-HCl, pH 8.3, 10 mM KCl and 2.5 mM MgCl<sub>2</sub>). Duplicate samples were denatured at 95°C for 5 min. and held at 72°C until 10 units of enzyme were added, and the samples were then cycled 0, 20, 25, 30, 35, 50 and 100 times as described in Example 3. The samples were analyzed on 1.2% agarose gels using ethidium bromide to visualize the presence of specific PCR product. The expected length of PCR product was about 800 bp. Reactions containing Taq DNA pol I, Stoffel and Tfl exo' fragments had visible product after 20 cycles, whereas reactions with Tfl holoenzyme showed product only after 30 cycles. The Tfl exo' fragment synthesized more product than the Stoffel fragment (FIGURE 7B). In general, there was some background, very likely because of the large amount of enzyme in the reaction. This background was not observed when 1 to 5 units of Tfl exo' fragment were used in a 35 cycle regimen. (FIGURE 7B, Right lane.) The polymerase activity in each tube was also determined as described above following completion of the PCR cycling, and the result plotted in a enzyme cycling stability curve (FIGURE 8). The half life was estimated to be: 25 cycles for both the Taq holoenzyme and Stoffel fragment, 20 cycles for the Tfl holoenzyme, and 16 cycles for the Tfl exo' fragment.

## EXAMPLE 11

### 20 DNA Sequencing with *T. flavus* DNA Polymerases

#### A.

Native and recombinant Tfl holoenzyme, Tfl exo' fragment, AmpliTaq, and Stoffel fragment were employed in the SEQUAL™ DNA Polymerase Sequencing System (CHIMERx) to test their performance in DNA sequencing using ssDNA template and labeled primer.

The primer FSP (Table 2) was end-labeled with T4 kinase and [<sup>32</sup>P]ATP according to the CHIMERx protocol. A 10 μl labeling reaction was prepared containing 0.5 μl primer (10 pmol/μl), 1.0 μl T4 Kinase 10X buffer (500 mM Tris-HCl pH 7.5, 100 mM MgCl<sub>2</sub>, 50 mM DTT, 1 mM



- 67 -

spermidine), 3.0  $\mu$ l [ $\gamma$ - $^{32}$ P] ATP (6000 Ci/mmol, 10  $\mu$ Ci/ $\mu$ l), 0.5  $\mu$ l T4 kinase (15 U/ $\mu$ l), and 5.0  $\mu$ l H<sub>2</sub>O. The labeling reaction was incubated at 37°C for 10 min., and the kinase was inactivated by incubation at 65°C for 10 min.

The sequencing reactions for native and recombinant Tfl holoenzyme and the exo' fragment were set up using 2-5 units of enzyme, according to CHIMERx conditions. Briefly, a reaction cocktail was prepared containing 16.0  $\mu$ l ssM13mp18 DNA (approx. 1  $\mu$ g), 5.0  $\mu$ l Sequel 5x buffer (250mM Tris-HCl, pH 9.5, 12.5 mM MgCl<sub>2</sub>), 1.0  $\mu$ l labeled primer (0.5 pmol/ $\mu$ l), and 2-5 units of enzyme. Four d/ddNTP mixtures were also prepared (A mix: 20  $\mu$ M dATP, 60  $\mu$ M dCTP, 60  $\mu$ M dGTP, 60  $\mu$ M dTTP, 300  $\mu$ M ddATP; C mix: 60  $\mu$ M dATP, 20  $\mu$ M dCTP, 60  $\mu$ M dGTP, 60  $\mu$ M dTTP, 150  $\mu$ M ddCTP; G mix: 60  $\mu$ M dATP, 60  $\mu$ M dCTP, 20  $\mu$ M dGTP, 60  $\mu$ M dTTP, 30  $\mu$ M ddGTP; T mix: 60  $\mu$ M dATP, 60  $\mu$ M dCTP, 60  $\mu$ M dGTP, 20  $\mu$ M dTTP, 400  $\mu$ M ddTTP). The sequencing reactions were performed by mixing 5  $\mu$ l of reaction cocktail with 1  $\mu$ l of the appropriate d/ddNTP mixture and heating the reaction tube at 65°C for 10 min. After this incubation 3  $\mu$ l of stop solution (EDTA/DTT/Bromophenol blue/xylene cyanol) were added and the reactions were placed on ice.

AmpliTaq reaction cocktail was prepared by using the 10x Reaction Buffer provided with the Cycle Sequencing Kit, which contains 2 units of enzyme in a final volume of 20  $\mu$ l. Stoffel fragment reaction cocktail (20  $\mu$ l) contained 4  $\mu$ l of 5x Stoffel fragment reaction buffer, 2  $\mu$ l of 25 mM MgCl<sub>2</sub>, both provided with the enzyme (Perkin Elmer), and 2 units of enzyme. Both cocktails included 1  $\mu$ g of ssDNA template and 1  $\mu$ l of labeled primer FSP. For both enzymes, sequencing reactions were performed by mixing 5  $\mu$ l reaction cocktail with 1  $\mu$ l d/ddNTP mixtures and incubating for 10 min. at 65°C. Three microliters of stop solution were then added to the reactions, and the reactions were placed on ice.

The reactions were heated at 90°C for 5 min. just before loading onto a 6% sequencing gel. One microliter of each sample was loaded

- 68 -

and electrophoresed at 3000 volts for 1.5 hours. The gel was autoradiographed and analyzed. FIGURE 10A photographically depicts a portion of a sequencing gel showing the same DNA sequence for all enzymes used, except for the native *T. flavus* DNA pol I holoenzyme control. Very little background was observed when the Tfl exo' fragment, Tfl holoenzyme and AmpliTaq were used. The Stoffel fragment had more ghost bands than the other enzymes. However, no attempt had been made to optimize the reaction conditions for the Stoffel fragment.

## B.

To demonstrate the utility of the recombinant Tfl holoenzyme and the Tfl exo' fragment in cycle sequencing with single-stranded DNA template, these enzymes were substituted into the SEQUAL™ DNA Polymerase and Cycle- SEQUAL™ Sequencing System (CHIMERx) and the protocols provided were followed.

The labeling protocol described above was repeated to create end-labeled primer. A 22  $\mu$ l reaction cocktail was then prepared containing approx. 20 ng ssM13mp18 DNA, 5.0  $\mu$ l 5X Sequal sequencing buffer, 1.0  $\mu$ l labeled primer (0.5 pmol/ $\mu$ l), balance H<sub>2</sub>O. Native or recombinant Tfl holoenzyme or Tfl exo' fragment was then added to the cocktail (0.5 units) and gently mixed. For comparison purposes, two units of AmpliTaq or Stoffel fragment were added to the ssM13mp18 DNA template (20 ng); the manufacturer's reaction conditions for the Perkin Elmer Cycle Sequencing Kit were followed for AmpliTaq, and, for the Stoffel fragment, 4  $\mu$ l of the Stoffel buffer and 2  $\mu$ l of the MgCl<sub>2</sub> solution provided with the enzyme were used.

The sequencing reactions were performed by mixing 5  $\mu$ l of reaction cocktail with 1  $\mu$ l of each d/ddNTP mixture (in separate tubes), adding a drop (~10 $\mu$ l) of mineral oil to each tube, and placing the tubes into a preheated (94°C) thermal cycler programmed to run the following cycle twenty times: 94°C for 15 seconds (denaturation), 70°C for 60 seconds

- 69 -

(extension). The reactions were cooled to 4°C after 20 cycles until 4 µl stop solution were added, and then the reactions were set on ice.

Immediately after heating the reaction mixtures at 90°C for 5 min., one microliter of each reaction mixture was loaded onto a 6% sequencing gel. FIGURE 10B shows that the Tfl exo' fragment and recombinant Tfl holoenzyme yield clean sequence data, whereas in the AmpliTaq lanes some ghost bands were observed. The Stoffel fragment, under the conditions used here, did not produce comparable sequencing data.

### C.

The utility of recombinant Tfl holoenzyme and Tfl exo' fragment for sequencing with internal labeling using double-stranded DNA template was demonstrated in a sequencing reaction in which a [ $\alpha^{35}$ S]-dATP labeling protocol and double stranded pUC19 template were used. The experiment was performed as outlined in the SEQUAL™ DNA Polymerase Sequencing System (CHIMERx) with 2 µg of pUC19 dsDNA and 2.5 units of the enzymes.

To promote efficient priming, the pUC19 double-stranded DNA template was denatured by adding deionized H<sub>2</sub>O to 18 µl, adding 2 µl of 2M NaOH, and incubating for 5 min. at room temperature. The reaction was neutralized by adding 2 µl of 2M ammonium acetate, pH 4.6, ethanol precipitated, air-dried, and resuspended in 10 µl deionized water.

For each enzyme, a 22.75 µl extension/labeling cocktail was prepared with the 2 µg denatured pUC19 dsDNA, 5.0 µl 5X Sequel buffer, 1.0 µl primer (0.5 pmol/µl), 1.0 µl alpha labeling mix (~45 µM each of dCTP, dGTP, dTTP), 0.25 µl [ $\alpha^{35}$ S] dATP (1000 Ci/mmol), 2.5 units enzyme, balance H<sub>2</sub>O. This cocktail was incubated at 65° for 10 min.

Extension/termination reactions were performed by adding 5 µl of extension/labeling cocktail to tubes containing 1 µl of the appropriate d/ddNTP mix, and mixing gently. The reaction tubes were immediately

- 70 -

placed at 65°C for 4 min., stopped by addition of 4  $\mu$ l step solution, and set on ice.

Each reaction was heated at 90° for 5 min. immediately before loading 1-2  $\mu$ l onto a sequencing gel. Results are depicted in FIGURE 10C.

- 5 Native Tfl holoenzyme (not shown) was compared to recombinant holoenzyme and to the Tfl *exo'* fragment. The bands were comparable for the holoenzymes. The quality of the sequence data is comparable although the signal was weaker when the Tfl *exo'* fragment was used.

#### D.

- 10 The utility of recombinant Tfl holoenzyme and the *exo'* fragment for double-stranded sequencing using a labeled sequencing primer was demonstrated by substitution of these enzymes into the SEQUAL™ System which uses 2  $\mu$ g pUC19 dsDNA and [ $\gamma^{32}$ P]-labeled primer FSP. The reactions were performed according to CHIMERx's protocol. More  
15 particularly, the double-stranded template was first denatured as described above, and then sequencing reactions were performed essentially as described in part A (substituting the pUC19 denatured dsDNA for ssM13mp18 template). As can be seen in FIGURE 10D, both the Tfl holoenzyme and the Tfl *exo'* fragment produced good sequence data.

20

### EXAMPLE 12

#### Polymerase Chain Reaction

- The utility of recombinant Tfl holoenzyme and the *exo'* fragment in PCR was demonstrated as follows. In a 0.5 ml reaction tube 85  $\mu$ l water, 2  $\mu$ l 10 mM dNTPs, 10  $\mu$ l 10 x Tfl Polymerase Reaction Buffer (10  
25 x buffer is 500 mM Tris-HCl, pH 9.0, 200 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 15 mM MgCl<sub>2</sub>), 1  $\mu$ l each of 50  $\mu$ M primers FTFL11 and RTFL12 (primer set 11-12), 50  $\mu$ l mineral oil and 1  $\mu$ l of 15  $\mu$ g/ml *T.flavus* genomic DNA were combined. After the initial denaturation step (Step 1), 5.5 and 11 units of Tfl *exo'* fragment, or 5 units of Tfl holoenzyme were added. As a control Taq pol I

- 71 -

in 1 x Taq Polymerase Reaction Buffer (Example 10) was used to amplify the genomic DNA. Amplifications were performed in a MJ Research PTC-100 Cyclor with external sensor control. The amplification program was: Step 1: 95°C for 5 min.; Step 2: hold at 72°C; Step 3: 55°C for 45 sec.; Step 4: 72°C for 5 min.; Step 5: 95°C for 15 sec.; Step 6: repeat steps 3-5 thirty-four times; Step 7: 55°C for 45 sec.; Step 8: 72°C for 20 min.; Step 9: hold at 4°C.

The amplification products were separated on 1.2% agarose gels. Primer set 11-12 gave a single amplification product from *T. flavus* genomic DNA. Five units of the Tfl exo<sup>-</sup> fragment produced a single product: the yield was slightly less than that obtained with Taq polymerase and better than the yield from Tfl holoenzyme.

### EXAMPLE 13

#### Thermal Cycle Labeling With Tfl DNA Pol I

The protocol described in Example 4 was used to demonstrate the utility of recombinant Tfl DNA pol I holoenzyme and the Tfl exo<sup>-</sup> fragment for thermal cycle labeling (TCL). See co-owned, co-pending U.S. Patent Application Serial No. 08/217,459, filed March 24, 1994, entitled "Materials and Methods for Restriction Endonuclease Applications." PCT Application No. US94/03246, filed March 24, 1994.

Thermal Cycle Labeling (TCL) is a procedure for labeling double-stranded DNA while simultaneously amplifying large amounts of the labeled probe. TCL of DNA requires two general steps: 1) generation of the sequence-specific oligonucleotides by CviII<sup>+</sup> (Molecular Biology Resources, Milwaukee, WI) restriction of the template DNA; and 2) repeated cycles of denaturation, annealing, and extension in the presence of a thermostable DNA polymerase or a functional fragment thereof which maintains polymerase activity. Optimal results are obtained after 20 such cycles, which is best performed in an automated thermal cycling instrument such as a Perkin-Elmer

- 72 -

Model 480 thermocycler. In conjunction with such an instrument, about 1.5 hr. is required to complete this protocol. If a thermal cycler is not available these reactions may be performed using heat blocks. As few as 5 cycles may yield probes with acceptable detection sensitivities. The generation of sequence specific oligonucleotides for use in this method may also be accomplished using the restriction endonuclease reagent CGase I (Molecular Biology Resources) or the restriction endonuclease Aci I which has as a recognition sequence CCGC.

Non-radioactive labeling of DNA using TCL is accomplished by mixing: 10 pg - 100 ng linearized template, 50 ng *Cvi*II<sup>+</sup>-digested primers, 1.5  $\mu$ l 10X labeling buffer, 2.5 - 5 units thermostable DNA polymerase, 1  $\mu$ l of 1mM Biotin-11-dUTP (Enzo Diagnostics, New York, New York), 1.5  $\mu$ l each of dATP, dCTP, and dGTP (2 mM), and 1.0  $\mu$ l 2mM dTTP. The reaction mixture is brought to a volume of 15  $\mu$ l with deionized H<sub>2</sub>O, overlaid with mineral oil and cycled through 20 rounds of denaturation, annealing and extension. A typical cycling regimen employs 20 cycles of denaturation at 91°C for 5 sec, annealing at 50°C for 5 sec and extension at 72°C for 30 sec. The reaction is then terminated by adding 1  $\mu$ l of 0.5M EDTA, pH 8.0. The amplified, labeled probe is a very heterogeneous mixture of fragments, which appears as a smear when analyzed by agarose gel electrophoresis.

The performance of recombinant Tfl DNA pol I holoenzyme, Tfl *exo*<sup>-</sup> fragment, Taq holoenzyme, and the Stoffel fragment (control) was assayed by substitution of these enzymes for the enzyme provided with the CHIMERx TCL kit (ZEPTO™ Labeling Kit). Five units of each enzyme and biotin-11-dUTP as the label were used. The substrate was pUC19 DNA.

After cycling of the samples the relative efficiency of the labeling reaction was determined by electrophoresis on a 0.7% agarose gel. The ethidium bromide gel staining of amplified DNA shows the characteristic smear for all enzymes used. The efficiency of incorporation was then determined by dot blot analysis.

- 73 -

The hybridized and developed filter showed that the holoenzymes (native and recombinant Tfl, and Taq) can be diluted 1: 10<sup>6</sup> and still generate a visible dot. The samples which were labeled with Tfl exo<sup>-</sup> or the Stoffel fragment can clearly be seen after a 1: 10<sup>4</sup> dilution. The 1: 10<sup>5</sup> dilution gave a weak signal when the exo<sup>-</sup> fragment was used for the TCL reaction.

Another aspect of the present invention involves a variation of TCL called Universal Thermal Cycle Labeling (UTCL) in which the extension primers are not supplied by CviJI\* restriction. Without intending to be limited to a particular theory, one explanation for the mechanism of UTCL is that the Tfl DNA pol I holoenzyme itself may supply endogenous "random" primers for enzymatic extension in a TCL-type reaction. Alternatively, some other explanation accounts for the mechanism of UTCL.

In a UTCL reaction, recombinant Tfl DNA pol I holoenzyme is combined with intact DNA template and is subjected to repeated cycles of denaturation, annealing, and extension. A radioactive- or non-isotopically-labeled deoxynucleotide triphosphate is incorporated during the extension step for subsequent detection purposes. The amplified, labeled probe represents a very heterogeneous mixture of fragments, which appears as a large molecular weight smear when analyzed by agarose gel electrophoresis. The utility of recombinant Tfl DNA pol I for Universal Thermal Cycle Labeling is demonstrated by substituting this enzyme in the UTCL protocol described in co-owned, copending U.S. Patent App. Ser. No. 08/217,459, filed March 24, 1994 (Example 12), incorporated herein by reference.

#### EXAMPLE 14

##### Reverse Transcription with *T. flavus* DNA Pol I Holoenzyme and exo<sup>-</sup> Fragment

RNA-dependent DNA polymerase activity of the Tfl DNA polymerases was analyzed using the following procedure: In a 0.5 ml reaction

- 74 -

tube, 2.5  $\mu$ l 1 M Tris-HCl, pH 8.3, 5  $\mu$ l of 0.6 M KCl, 5  $\mu$ l of 0.04 M  $MgCl_2$ , 17.5  $\mu$ l of water, 10  $\mu$ l of 2 mM poly rA:dT (the substrate) and 5  $\mu$ l 5 mM [ $\alpha$ - $^{32}P$ ]TTP at 10  $\mu$ Ci/ml were combined. After incubation at 55°C for 5 min., the reaction was started by the addition of 5  $\mu$ l of enzyme (Five DNA-dependent DNA polymerase units per  $\mu$ l). The reaction was allowed to proceed at 55°C for 30 min., and terminated by taking a 40  $\mu$ l aliquot and adding it to 50  $\mu$ l of 10% tRNA, 2% sodium pyrophosphate. The samples were precipitated with TCA and the enzyme activity was determined as described above. The RNA-dependent polymerase activity of the native and the recombinant Tfl DNA pol I was determined to be about 6% of the DNA-dependent polymerase activity. When 10 (RNA-dependent DNA polymerase) units of AMV-RT (Molecular Biology Resources, Inc., Cat. No. 1372-01) were compared to 10 (DNA-dependent DNA polymerase) units of Tfl DNA pol I it was found that nTfl DNA pol I possess 2.4% and rTfl DNA pol I 1.6% of the RNA-dependent DNA polymerase activity of AMV-RT. Titration of the  $MgCl_2$  and the  $MnCl_2$  concentration revealed that the native and the recombinant holoenzymes prefer  $MgCl_2$  over  $MnCl_2$  for RT activity.

The Tfl exo' fragment has a lower RT activity than the holoenzyme, but has a broader temperature range for activity. First strand cDNA synthesis with the holoenzymes apparently yields a product of the same length as that obtained by using AMV-RT. The recombinant *T. flavus* DNA polymerase I and the exo' fragment both exhibit reverse transcriptase function which can be used in applications such as RT-PCR or cDNA preparation at elevated temperatures.

#### EXAMPLE 15

##### Comparison of the Processivity of DNA Polymerase Enzymes

Using a modification of a procedure described by Tabor et al., *J. Biol. Chem.* 262: 16212-16223 (1987), the processivity of native and



- 75 -

recombinant Tfl DNA pol I holoenzyme, Tfl *exo'* fragment, and Taq DNA pol I holoenzyme were compared. The "processivity" of a DNA polymerase enzyme is a measure of the rate at which the enzyme moves forward along a template while catalyzing DNA synthesis, i.e., a measure of the speed at which DNA polymerization takes place in the presence of the enzyme.

To prepare the assay, a 60  $\mu$ l reaction cocktail was prepared with 3  $\mu$ g M13 mp18 ssDNA, 12  $\mu$ l ddATP mix (20  $\mu$ M dATP; 60  $\mu$ M each of dCTP, dGTP, and dTTP; 300  $\mu$ M ddATP), 3.0  $\mu$ l  $\alpha$ -<sup>32</sup>P labeled forward sequencing primer (3  $\mu$ g/ $\mu$ l), 12  $\mu$ l 5x reaction buffer (250 mM Tris-HCl, pH 9.5; 12.5 mM MgCl<sub>2</sub>), balance H<sub>2</sub>O. Additionally, dilutions of native and recombinant Tfl holoenzyme, Tfl *exo'* fragment, and Taq holoenzyme were prepared with appropriate storage buffer to create enzyme solutions of 0.125 and 0.0125 units/ $\mu$ l for the holoenzymes and 0.25 and 0.025 units/ml for Tfl *exo'* fragment.

To perform the assay, 7.0  $\mu$ l of the reaction cocktail were mixed with 2.0  $\mu$ l of diluted DNA polymerase enzyme. By using 0.25 and 0.025 units of Taq, nTfl, and rTfl holoenzyme and 0.5 and 0.05 units of *exo'* fragment per reaction, reactions containing approximately 1:100 and 1:1000 enzyme molecule: template molecule are obtained. The use of such low polymerase concentrations minimizes the "bumping off" from template by competing polymerase molecules. Reaction mixtures were incubated at 65°C and 3  $\mu$ l samples were removed at 1.0, 2.5 and 6.0 minute time points. Reactions were stopped by adding 1.0  $\mu$ l stop buffer (EDTA/DTT/BromoPhenol Blue/xylene cyanol), were heated at 90°C for 3 min., and were loaded onto 7.5% polyacrylamide sequencing gels. The gels were electrophoresed until the bromophenol blue dye was about 3/4 down the gel, and an autoradiograph of the gel was taken overnight at -70°C.

With this assay, a highly processive enzyme produces strong, slow-mobility (larger) labeled bands on an autoradiograph, whereas a less processive DNA polymerase produces higher-mobility (smaller) fragments

- 76 -

and/or bands with less intensity. Autoradiographs of the 6 min. incubation/1:100 enzyme:template reactions revealed the *exo'* fragment produced bands with the most intensity, followed by the *rTfl* and *nTfl* holoenzyme, then the *Taq* holoenzyme. The length of fragments obtained by the four enzymes was very comparable. Autoradiographs from the 1:1000 enzyme:template reaction indicate that processivity (from best to least) is *Tfl* *exo'* fragment > *Taq* holoenzyme > *nTfl* holoenzyme and *rTfl* holoenzyme. These results indicate that *Tfl* *exo'* fragment has greater processivity than either *Tfl* holoenzyme (native or recombinant) or *Taq* holoenzyme.

#### EXAMPLE 16

##### Large-Scale Purification of Recombinant *Tfl* DNA Polymerase I Holoenzyme and *Exo'* Fragment

Both the recombinant *Tfl* holoenzyme and *Tfl* *exo'* fragment were purified on a large "production" scale by modifying the procedure described above for purifying native *Tfl* holoenzyme.

Four hundred sixty grams of induced *E. coli* DH5 $\alpha$ F'IQ cells transformed with pTFLRT4 (cultured and frozen as described above) were thawed and suspended in 2500 ml of lysis buffer A (20 mM Tris-HCl, pH 8; 0.5 mM EDTA; 7 mM  $\beta$ -mercaptoethanol; 10 mM MgCl<sub>2</sub>). For *Tfl* *exo'* fragment, 787 grams of *E. coli* transformed with p21EHcM1.1 (cultured and frozen as described above) were used. Phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 0.3 mM.

The suspension was then treated with 0.2 g/l of lysozyme (predissolved in lysis buffer) at 4°C for 1 hr. Cells were homogenized twice at 9000 psi in a Manton Gaulin homogenizer, with the suspension chilled to approximately 10°C between passes. New PMSF was added to 0.2 g/l before, between and after passes. The suspension of lysed cells was divided into 300 ml portions, heated to 65°C for 1 hr., cooled down to 4°C, and centrifuged for 30 min. at 13,500 x g.

- 77 -

Following the centrifugation, NaCl and polyethyleneimine (PEI) (10% w/v, pH 7.0) were added to the heat-treated supernatant to a final concentration of 0.5 M and to 0.1%, respectively. The sample was mixed well and centrifuged at 13,500 x g for 1 hour.

5           The supernatant from the twice-centrifuged, heat-treated lysate was desalted by diluting with 10 liters of DE52 column buffer (20 mM Tris-HCl, pH 8.0; 0.5 mM EDTA; 7 mM  $\beta$ ME) and concentrated to approximately 4 liters using an Amicon S10Y30 Spiral Ultrafiltration cartridge. The dilution/concentration step was repeated two times, with a  
10          final concentrated volume of about 4 liters.

          The desalted sample was batch contacted with 400 g of equilibrated Whatman DE52 ion exchange resin (Maidstone, England). The suspension was collected on a sintered glass funnel and washed 3 times with 1 volume of DE52 column buffer. The resin was then resuspended in a  
15          minimal volume of buffer and poured into a column (4.5 x 50 cm), packed and washed with an additional volume of buffer. The column was eluted with a 0-0.5 M NaCl linear gradient (total gradient volume: 2000 ml). Twenty-five ml fractions were collected at a rate of about 5 ml/min. Peak fractions (fractions containing DNA polymerase activity) were determined by a  
20          modified DNA polymerase assay described by Kaledin et al., *Biokhimiya* 45:644-651 (1980), pooled and dialyzed in approximately twenty-five volumes of Affi-Gel Blue (AGB) column buffer (20 mM Tris-HCl, pH 7.5; 0.5 mM EDTA; 10 mM  $\beta$ ME; 10 mM  $MgCl_2$ ; 0.02% Brij 35).

          The dialyzed DE52 peak fractions were applied to an AGB  
25          column (4.4 x 40 cm, 600 ml packed volume, MBR Blue, Molecular Biology Resources, Milwaukee WI), which was washed with 2 column volumes of AGB column buffer, and eluted with a 0-1.2 M NaCl linear gradient (total gradient volume: 2000 ml). To elute the exo' fragment, a 0-1.5 M NaCl linear gradient was employed. Twenty-five ml fractions were collected at a

- 78 -

rate of 1-5 ml/min. The peak fractions were dialyzed as above in AGB buffer.

The dialyzed AGB peak fractions were applied to a Heparin Agarose column (4.4 x 16.5 cm, 250 ml packed volume (Bio-Rad Affigel Heparin or Heparin Agarose from Molecular Chimetrics, Madison, WI)), which was washed with approximately 2 column volumes (until effluent is no longer colored, and column resin is white in appearance), and eluted with a 0.1-1.0 M NaCl linear gradient (total gradient volume: 1500 ml). To elute the exo' fragment, a 0.15-1.0 M NaCl linear gradient was employed. Twenty-five ml fractions were collected at a rate of 1-5 ml/min. The peak fractions were dialyzed in HP Q Sepharose Column Buffer (20 mM Tris-HCl, pH 7.5; 0.5 mM EDTA; 7 mM  $\beta$ ME; 0.1% Brij 35).

The dialyzed heparin agarose peak fractions were filtered through a 0.2  $\mu$ m filter and applied at 4 ml/min. to the HP Q Sepharose column (Pharmacia, Uppsala, Sweden) on FPLC. The column was washed with several column volumes of buffer, and eluted with a 0-0.25 M NaCl linear gradient. Ten ml fractions were collected at 4 ml/minute. The peak fractions were dialyzed in HP S Column Buffer (20 mM Na-Citrate, pH 6.0; 1 mM EDTA; 7 mM  $\beta$ ME; 0.1% Brij 35) or diluted in the same buffer, depending on the volume of the fraction pool.

The dialyzed (or diluted) HP Q peak fractions were filtered through a 0.2  $\mu$ m filter and the HP S column (Pharmacia) was run as above, washing with HP S Column buffer and eluting with a 0-0.25 M NaCl gradient. Peak fractions were pooled and dialyzed against 4 liters of Final Storage Buffer (50 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 5 mM DTT, 50% glycerol). The final product was diluted to a concentration of 5000 U/ml in the above buffer including 0.5 % Tween 20 (Sigma Chemical Co., St. Louis, MO) and 0.5 % Nonidet P40 (Fluka Biochemika, Buchs, Switzerland) as stabilizers and stored at -20°C. To purify the recombinant holoenzyme, the HP S column purification was unnecessary, and therefore was omitted.

- 79 -

Samples from the above-described preparations were electrophoresed using SDS-Page and visualized with silver staining. The rTfl holoenzyme and exo<sup>-</sup> fragments appeared as single bands having apparent molecular weights of 88,000 and 63,000 kDa, respectively, each being greater than 95% pure. A quantitative analysis of the enzymes prepared using the above-described purification procedure is as shown in Table 6:

TABLE 6			
Enzyme	Quantity of Cells	Specific Activity (Units/mg protein)	Yield (Units/g cells)
nTfl Holo (Example 1)	1200 g	50,000 U/mg	1,700 U/g
rTfl Holo	460 g	70,000 U/mg	4,300 U/g
rTfl exo <sup>-</sup>	787 g	192,000 U/mg	5,600 U/g

The biological activities of the recombinant enzymes purified by the above-described protocol were analyzed using the assays described in preceding Examples. In the endonuclease activity assay described in Example 9, five, ten, and twenty unit challenges resulted in less than 5% conversion of supercoiled pBR322 to the linear form. The results of other assays described in Example 9 are summarized in Table 7:

TABLE 7		
ASSAY	rTfl holo	Tfl Exo <sup>-</sup>
ds DNase	0% slope/unit	0% slope/unit
ss DNase	0% slope/unit	0% slope/unit
3' Exonuclease	0% slope/unit	0.06% slope/unit
5' Exonuclease	0.48% slope/unit	0% slope/unit

- 80 -

**Deposit of Biological Materials:** The following plasmids have been deposited with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville MD 20852 (USA) pursuant to the provisions of the Budapest Treaty:

5	Designation	Deposit Date	ATCC No.	Host Strain
	pTFLRT4	May 26, 1994	69633	DH5 $\alpha$ F'IQ
	P21EHcM1.1	May 26, 1994	69632	DH5 $\alpha$ F'

Availability of the deposited materials is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

The present invention has been described with reference to specific examples and embodiments. However, this application is intended to cover those changes and substitutions which are apparent and may be made by those skilled in the art without departing from the spirit and scope of the claims.

- 81 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT:
  - (A) NAME: Molecular Biology Resources, Inc.
  - (B) STREET: 5520 W. Burleigh Street
  - (C) CITY: Milwaukee
  - (D) STATE: Wisconsin
  - (E) COUNTRY: United States of America
  - (F) POSTAL CODE: 53210
- (ii) TITLE OF INVENTION: Biologically Active Fragments of  
Thermus Flavus DNA Polymerase
- (iii) NUMBER OF SEQUENCES: 51
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
  - (B) STREET: 6300 Sears Tower, 233 South Wacker Drive
  - (C) CITY: Chicago
  - (D) STATE: Illinois
  - (E) COUNTRY: United States of America
  - (F) POSTAL CODE: 60606-6402
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Gass, David A.
  - (B) REGISTRATION NUMBER: 38,153
  - (C) REFERENCE/DOCKET NUMBER: 28003/31716
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 312/474-6300
  - (B) TELEFAX: 312/474-0448
  - (C) TELEX: 25-3856

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3048 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 301..2805

- 82 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TACTTCGGCG GGGTGAAGCT CGGGGCCGGG GGGCTTGTGC GGGCCTACGG GGGGGTGGCG	60
CGCGAGGCCT TAAGCGGGCG CCCAAGGTCC CCTTGGTGGA GCGGGTGGGG CTCGCTTCC	120
TGCTGCCCTT CGCCGAGGTG GGGCGGTCT ACGCCCTCCT GGAGGCCCGC GCCCTGAAGG	180
CGGAGGAGAC CTACACCCCG GAGGGCGTGC GCTTCGCCCT CCTCTCCCC AAGCCCGAGC	240
GGGAAGGTTT CCTCAGGGCG CTCTGGACG CCACCCGGGG ACAGGTGGCC CTGGAGTAGC	300
ATG GAG GCG ATC GTT CCG CTC TTT GAA CCC AAA GGC CGG GTC CTC CTG	348
Met Glu Ala Ile Val Pro Leu Phe Glu Lys Gly Arg Val Leu Leu	
1 5 10 15	
GTG GAC GGC CAC CAC CTG GCC TAC CGC ACC TTC TTC GCC CTG AAG GGC	396
Val Asp Gly His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly	
20 25 30	
CTC ACC ACG AGC CGG GGC GAA CCG GTG CAG GCG GTC TAC GGC TTC GCC	444
Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala	
35 40 45	
AAG AGC CTC CTC AAG GCC CTG AAG GAG GAC GGG TAC AAG GCC GTC TTC	492
Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe	
50 55 60	
GTG GTC TTT GAC GCC AAG GCC CCC TCC TTC CGC CAC GAG GCC TAC GAG	540
Val Val Phe Asp Ala Val Ala Pro Ser Phe Arg His Glu Ala Tyr Glu	
65 70 75 80	
GCC TAC AAG GCG GGG AGG GCC CCG ACC CCC GAG GAC TTC CCC CGG CAG	588
Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln	
85 90 95	
CTC GCC CTC ATC AAG GAG CTG GTG GAC CTC CTG GGG TTT ACC CGC CTC	636
Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu	
100 105 110	
GAG GTC CCC GGC TAC GAG GCG GAC GAC GTC CTC GCC ACC CTG GCC AAG	684
Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys	
115 120 125	
AAG GCG GAA AAG GAG GGG TAC GAG GTG CGC ATC CTC ACC GCC GAC CGC	732
Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg	
130 135 140	
GAC CTC TAC CAA CTC GTC TCC GAC CGC GTC GTC GTC CTC CAC CCC GAG	780
Asp Leu Tyr Gln Leu Val Ser Asp Arg Val Val Val Leu His Pro Glu	
145 150 155 160	
GGC CAC CTC ATC ACC CCG GAG TGG CTT TGG GAG AAG TAC GGC CTC AAG	828
Gly His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Lys	
165 170 175	
CCG GAG CAG TGG GTG GAC TTC CGC GCC CTC GTG GGG GAC CCC TCC GAC	876
Pro Glu Gln Trp Val Asp Phe Arg Ala Leu Val Gly Asp Pro Ser Asp	
180 185 190	



- 83 -

AAC	CTC	CCC	GGG	GTC	AAG	GGC	ATC	GGG	GAG	AAG	ACC	GCC	CTC	AAG	CTC	924
Asn	Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	
		195					200					205				
CTC	AAG	GAG	TGG	GGA	AGC	CTG	GAA	AAC	CTC	CTC	AAG	AAC	CTG	GAC	CGG	972
Leu	Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	
		210				215					220					
GTA	AAG	CCA	GAA	AAC	GTC	CGG	GAG	AAG	ATC	AAG	GCC	CAC	CTG	GAA	GAC	1020
Val	Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	
		225			230					235				240		
CTC	AGG	CTT	TCC	TTG	GAG	CTC	TCC	CGG	GTG	CGC	ACC	GAC	CTC	CCC	CTG	1068
Leu	Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	
				245					250					255		
GAG	GTG	GAC	CTC	GCC	CAG	GGG	CGG	GAG	CCC	GAC	CGG	GAG	GGG	CTT	AGG	1116
Glu	Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	
			260					265					270			
GCC	TTC	CTG	GAG	AGG	CTG	GAG	TTC	GGC	AGC	CTC	CTC	CAC	GAG	TTC	GGC	1164
Ala	Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	
		275					280					285				
CTC	CTG	GAG	GCC	CCC	GCC	CCC	CTG	GAG	GAG	GCC	CCC	TGG	CCC	CCG	CCG	1212
Leu	Leu	Glu	Ala	Pro	Ala	Pro	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	
		290				295					300					
GAA	GGG	GCC	TTC	GTG	GGC	TTC	GTC	CTC	TCC	CGC	CCC	GAG	CCC	ATG	TGG	1260
Glu	Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Pro	Glu	Pro	Met	Trp	
		305			310					315				320		
GCG	GAG	CTT	AAA	GCC	CTG	GCC	GCC	TGC	AGG	GAC	GSC	CGG	GTG	CAC	CGG	1308
Ala	Glu	Leu	Lys	Ala	Leu	Ala	Ala	Cys	Arg	Asp	Gly	Arg	Val	His	Arg	
				325					330					335		
GCA	GCA	GAC	CCC	TTG	GCG	GGG	CTA	AAG	GAC	CTC	AAG	GAG	GTG	CGG	GGT	1356
Ala	Ala	Asp	Pro	Leu	Ala	Gly	Leu	Lys	Asp	Leu	Lys	Glu	Val	Arg	Gly	
			340					345					350			
CTC	CTC	GCC	AAG	GAC	CTC	GCC	GTC	TTG	GCC	TCG	AGG	GAG	GGG	CTA	GAC	1404
Leu	Leu	Ala	Lys	Asp	Leu	Ala	Val	Leu	Ala	Ser	Arg	Glu	Gly	Leu	Asp	
		355					360					365				
CTC	GTG	CCC	GGG	GAC	GAC	CCC	ATG	CTC	CTC	GCC	TAC	CTC	CTG	GAC	CCC	1452
Leu	Val	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	
		370				375					380					
TCC	AAC	ACC	ACC	CCC	GAG	GGG	GTG	GCG	CGG	CGC	TAC	GGG	GGG	GAG	TGG	1500
Ser	Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	
		385			390					395				400		
ACG	GAG	GAC	GCC	GCC	CAC	CGG	GCC	CTC	CTC	TCG	GAG	AGG	CTC	CAT	CGG	1548
Thr	Glu	Asp	Ala	Ala	His	Arg	Ala	Leu	Leu	Ser	Glu	Arg	Leu	His	Arg	
			405					410						415		
AAC	CTC	CTT	AAG	CGC	CTC	GAG	GGG	GAG	GAG	AAG	CTC	CTT	TGG	CTC	TAC	1596
Asn	Leu	Leu	Lys	Arg	Leu	Glu	Gly	Glu	Glu	Lys	Leu	Leu	Trp	Leu	Tyr	
			420				425						430			

- 84 -

CAC His	GAG Glu	GTG Val	GAA Glu	AAG Lys	CCC Pro	CTC Leu	TCC Ser	CGG Arg	GTC Val	CTG Leu	GCC Ala	CAC His	ATG Met	GAG Glu	GCC Ala	1644
435							440					445				
ACC Thr	GGG Gly	GTA Val	CGG Arg	CTG Leu	GAC Asp	GTG Val	GCC Ala	TAC Tyr	CTG Leu	CAG Gln	GCC Ala	CTT Leu	TCC Ser	CTG Leu	GAG Glu	1692
450						455					460					
CTT Leu	CGC Ala	GAG Glu	GAG Glu	ATC Ile	CGC Arg	CGC Arg	CTC Leu	GAG Glu	GAG Glu	GAG Val	GTC Phe	TTC Arg	CGC Arg	TTG Leu	CGC Ala	1740
465					470					475					480	
GGC Gly	CAC His	CCC Pro	TTC Phe	AAC Asn	CTC Leu	AAC Asn	TCC Ser	CGG Arg	GAC Gln	CAG Gln	CTG Leu	GAA Glu	AGG Arg	GTG Val	CTC Leu	1788
				485					490					495		
TTT Phe	GAC Asp	GAG Glu	CTT Leu	AGG Arg	CTT Leu	CCC Pro	GCC Ala	TTG Leu	GGG Gly	AAG Lys	ACG Thr	CAA Gln	AAG Lys	ACG Thr	GGC Gly	1836
			500					505					510			
AAG Lys	CGC Arg	TCC Ser	ACC Thr	AGC Ser	GCC Ala	GCG Ala	GTG Val	CTG Leu	GAG Glu	GCC Ala	CTA Leu	CGG Arg	GAG Glu	GCC Ala	CAC His	1884
		515					520					525				
CCC Pro	ATC Ile	GTG Val	GAG Glu	AAG Lys	ATC Ile	CTC Gln	CAG His	CGG Arg	GAG Glu	CTC Gln	ACC Thr	AAG Lys	CTC Lys	AAG Leu		1932
	530					535					540					
AAC Asn	ACC Thr	TAC Tyr	GTG Val	GAC Asp	CCC Pro	CTC Leu	CCA Pro	AGC Ser	CTC Leu	GTC Val	CAC His	CCG Pro	AGG Arg	ACG Thr	GGC Gly	1980
545					550					555					560	
CGC Arg	CTC Leu	CAC His	ACC Thr	CGC Arg	TTC Phe	AAC Asn	CAG Gln	ACG Thr	GCC Ala	ACG Thr	GCC Ala	ACG Thr	GGG Gly	AGG Arg	CTT Leu	2028
				565					570					575		
AGT Ser	AGC Ser	TCC Ser	GAC Asp	CCC Pro	AAC Asn	CTG Leu	CAG Gln	AAC Asn	ATC Ile	CCC Pro	GTC Val	CGC Arg	ACC Thr	CCC Pro	TTG Leu	2076
			580					585					590			
GGC Gly	CAG Gln	AGG Ile	ATC Arg	CGC Arg	CGG Arg	GCC Phe	TTC Val	GTG Ala	GCC Ala	GAG Glu	GCG Ala	GGA Gly	TGG Trp	CGC Ala	TTG Leu	2124
		595				600						605				
GTG Val	GCC Ala	CTG Leu	GAC Asp	TAT Tyr	AGC Ser	CAG Gln	ATA Ile	GAG Glu	CTC Leu	CGC Arg	GTC Val	CTC Ala	GCC Leu	CAC His	CTC Leu	2172
		610			615						620					
TCC Ser	GGG Gly	GAC Asp	GAG Glu	AAC Asn	CTG Leu	ATC Ile	AGG Arg	GTC Val	TTC Phe	CAG Gln	GAG Glu	GGG Gly	AAG Lys	GAC Asp	ATC Ile	2220
625					630					635				640		
CAC His	ACC Thr	CAG Gln	ACC Thr	GCA Ala	AGC Ser	TGG Trp	ATG Met	TTC Phe	GGC Gly	GTC Val	CCC Pro	CCG Pro	GAG Glu	GCC Ala	GTG Val	2268
				645					650					655		
GAC Asp	CCC Pro	CTG Leu	ATG Met	CGC Arg	CGG Arg	GCG Ala	GCC Ala	AAG Lys	ACG Thr	GTG Val	AAC Asn	TTC Phe	GGC Gly	GTC Val	CTC Leu	2316
				660				665					670			

- 85 -

TAC GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTT GCC ATC CCC TAC Tyr Gly Met Ser Ala His Arg Leu Ser Gln Leu Ala Ile Pro Tyr 675 680 685	2364
GAG GAG GCG GTG GCC TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys 690 695 700	2412
GTG CGG GCC TGG ATA GAA AAG ACC CTG GAG GAG GGG AGG AAG CGG GGC Val Arg Ala Trp Ile GAA AAG ACC CTG Leu Glu Glu Gly Arg Lys Arg Gly 705 710 715	2460
TAC GTG GAA ACC CTC TTC GGA AGA AGG CGC TAC GTG CCC GAC CTC AAC Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn 725 730 735	2508
GCC CGG GTG AAG AGC GTC AGG GAG GCC GCG GAG CGC ATG GCC TTC AAC Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn 740 745 750	2556
ATG CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTC GCC ATG GTG Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val 755 760 765	2604
AAG CTC TTC CCC CGC CTC CGG GAG ATG GGG GCC CGC ATG CTC CTC CAG Lys Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln 770 775 780	2652
GTC CAC GAC GAG CTC CTC CTG GAG GCC CCC CAA GCG CGG GCC GAG GAG Val His Asp Glu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu 785 790 795	2700
GTG GCG GCT TTG GCC AAG GAG GCC ATG GAG AAG GCC TAT CCC CTC GCC Val Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala 805 810 815	2748
GTG CCC CTG GAG GTG GAG GTG GGG ATG GGG GAG GAC TGG CTT TCC GCC Val Pro Leu Glu Val Glu Val Gly Met Gly Glu Asp Trp Leu Ser Ala 820 825 830	2796
AAG GGT TAGGGGGGCC CTGCGGTTTA GAGGAAGTTC AAGGGGTGT CCCTCAGAAA Lys Gly	2852
CGCCTCCAGG GGAACGCCCT CTGCGGCTAC CAGGAGGCCT TTAGCCCCAA AGGTGCGGGT	2912
GAAGGCTTCC AGGCCCTGGG TTCTTTTAAA GGGGCGGCTT TTGACCTCGA GGGCCAGGAG	2972
GCGCTTTCCC TTTGAAGGA CAAAGTCACT TCCTGGTCCC TTTCGCGCCA GTAGTACACC	3032
TCAAACCCCC CCTGGT	3048

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 834 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

- 86 -

Met Glu Ala Ile Val Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu  
 1 5 10 15  
 Val Asp Gly His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly  
 20 25 30  
 Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala  
 35 40 45  
 Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe  
 50 55 60  
 Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu  
 65 70 75 80  
 Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln  
 85 90 95  
 Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu  
 100 105 110  
 Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys  
 115 120 125  
 Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg  
 130 135 140  
 Asp Leu Tyr Gln Leu Val Ser Asp Arg Val Val Val Leu His Pro Glu  
 145 150 155 160  
 Gly His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Lys  
 165 170 175  
 Pro Glu Gln Trp Val Asp Phe Arg Ala Leu Val Gly Asp Pro Ser Asp  
 180 185 190  
 Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Leu Lys Leu  
 195 200 205  
 Leu Lys Glu Trp Gly Ser Leu Glu Asn Leu Leu Lys Asn Leu Asp Arg  
 210 215 220  
 Val Lys Pro Glu Asn Val Arg Glu Lys Ile Lys Ala His Leu Glu Asp  
 225 230 235 240  
 Leu Arg Leu Ser Leu Glu Leu Ser Arg Val Arg Thr Asp Leu Pro Leu  
 245 250 255  
 Glu Val Asp Leu Ala Gln Gly Arg Glu Pro Asp Arg Glu Gly Leu Arg  
 260 265 270  
 Ala Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly  
 275 280 285  
 Leu Leu Glu Ala Pro Ala Pro Leu Glu Glu Ala Pro Trp Pro Pro Pro  
 290 295 300  
 Glu Gly Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp  
 305 310 315 320  
 Ala Glu Leu Lys Ala Leu Ala Ala Cys Arg Asp Gly Arg Val His Arg  
 325 330 335

- 87 -

Ala Ala Asp Pro Leu Ala Gly Leu Lys Asp Leu Lys Glu Val Arg Gly  
 340 345 350  
 Leu Leu Ala Lys Asp Leu Ala Val Leu Ala Ser Arg Glu Gly Leu Asp  
 355 360 365  
 Leu Val Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro  
 370 375 380  
 Ser Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp  
 385 390 395 400  
 Thr Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg Leu His Arg  
 405 410 415  
 Asn Leu Leu Lys Arg Leu Glu Gly Glu Glu Lys Leu Leu Trp Leu Tyr  
 420 425 430  
 His Glu Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala  
 435 440 445  
 Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu  
 450 455 460  
 Leu Ala Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala  
 465 470 475 480  
 Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu  
 485 490 495  
 Phe Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly  
 500 505 510  
 Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His  
 515 520 525  
 Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys  
 530 535 540  
 Asn Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly  
 545 550 555 560  
 Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu  
 565 570 575  
 Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu  
 580 585 590  
 Gly Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu  
 595 600 605  
 Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu  
 610 615 620  
 Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile  
 625 630 635 640  
 His Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val  
 645 650 655  
 Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu  
 660 665 670

- 88 -

Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr  
 675 680 685  
 Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys  
 690 695 700  
 Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Gly Arg Lys Arg Gly  
 705 710 715 720  
 Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn  
 725 730 735  
 Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn  
 740 745 750  
 Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val  
 755 760 765  
 Lys Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln  
 770 775 780  
 Val His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu  
 785 790 795 800  
 Val Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala  
 805 810 815  
 Val Pro Leu Glu Val Glu Val Gly Met Gly Glu Asp Trp Leu Ser Ala  
 820 825 830  
 Lys Gly

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1794 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1794

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATG GAA GAC CTC AGG CTT TCC TTG GAG CTC TCC CGG GTG CGC ACC GAC	48
Met Glu Asp Leu Arg Leu Ser Leu Glu Leu Ser Arg Val Arg Thr Asp	
1 5 10 15	
CTC CCC CTG GAG GTG GAC CTC GCC CAG GGG CGG GAG CCC GAC CGG GAG	96
Leu Pro Leu Glu Val Asp Leu Ala Gln Gly Arg Glu Pro Asp Arg Glu	
20 25 30	
GGG CTT AGG GCC TTC CTG GAG AGG CTG GAG TTC GGC AGC CTC CTC CAC	144
Gly Leu Arg Ala Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His	
35 40 45	

- 89 -

GAG Glu	TTC Phe	GGC Gly	CTC Leu	CTG Leu	GAG Glu	GCC Ala	CCC Pro	GCC Pro	CCC Pro	CTG Leu	GAG Glu	GAG Glu	GCC Ala	CCC Ala	TGG Trp	192
50						55				60						
CCC Pro	CCG Pro	CCG Pro	GAA Gly	GGG Gly	GCC Ala	TTC Phe	GTG Val	GGC Gly	TTC Phe	GTG Val	CTC Leu	TCC Ser	CGC Arg	CCC Arg	GAG Glu	240
65					70					75					80	
CCC Pro	ATG Met	TGG Trp	GCG Ala	GAG Glu	CTT Leu	AAA Lys	GCC Ala	CTG Leu	CCC Ala	GCC Ala	TCC Cys	AGG Arg	GAC Asp	GGC Gly	CGG Arg	288
				85				90						95		
GTG Val	CAC His	CGG Arg	GCA Ala	GAC Ala	GAC Asp	CCC Pro	TTG Leu	GCG Ala	GGG Gly	CTA Leu	AAG Lys	GAC Asp	CTC Leu	AAG Lys	GAG Glu	336
			100					105						110		
GTG Val	CGG Arg	GGT Gly	CTC Leu	CTC Leu	GCC Ala	AAG Lys	GAC Asp	CTC Leu	GCC Ala	GTG Val	TTG Leu	GCC Ala	TCC Ser	AGG Arg	GAG Glu	384
		115					120					125				
GGG Gly	CTA Leu	GAC Asp	CTC Leu	GTG Val	CCC Pro	GGG Gly	GAC Asp	CCC Pro	ATG Met	CTC Leu	CTC Leu	GCC Ala	TAC Tyr	CTC Leu		432
	130					135				140						
CTG Leu	GAC Asp	CCC Pro	TCC Ser	AAC Asn	ACC Thr	ACC Thr	CCC Pro	GAG Glu	GGG Gly	GTG Val	CGG Ala	CGG Arg	CGC Arg	TAC Tyr	GGG Gly	480
	145				150					155					160	
GGG Gly	GAG Glu	TGG Trp	ACG Thr	GAG Glu	GAC Asp	GCC Ala	GCC Ala	CAC His	CGG Arg	GCC Ala	CTC Leu	CTC Leu	TCC Ser	GAG Glu	AGG Arg	528
				165				170						175		
CTC Leu	CAT His	CGG Arg	AAC Asn	CTC Leu	CTT Leu	AAG Lys	CGC Arg	CTC Leu	GAG Glu	GGG Gly	GAG Glu	GAG Glu	AAG Lys	CTC Leu	CTT Leu	576
			180					185					190			
TGG Trp	CTC Leu	TAC Tyr	CAC His	GAG Glu	GTG Val	GAA Glu	AAG Lys	CCC Pro	CTC Leu	TCC Ser	CGG Arg	GTG Val	CTG Leu	GCC Ala	CAC His	624
		195					200					205				
ATG Met	GAG Glu	GCC Ala	ACC Thr	GGG Gly	GTA Val	CGG Glu	CTG Leu	GAC Asp	GTG Val	GCC Ala	TAC Leu	CTG Gln	CAG Gln	GCC Ala	CTT Leu	672
	210					215					220					
TCC Ser	CTG Leu	GAG Glu	CTT Leu	GCG Ala	GAG Glu	GAG Ile	ATC Arg	CGC Arg	CGC Arg	CTC Leu	GAG Glu	GAG Glu	GAG Glu	GTG Val	TTC Phe	720
	225				230					235					240	
CGC Arg	TTG Leu	GCG Ala	GGC Gly	CAC His	CCC Pro	TTC Phe	AAC Asn	CTC Leu	AAC Leu	TCC Ser	CGG Arg	GAC Asp	CAG Gln	CTG Leu	GAA Glu	768
				245				250						255		
AGG Arg	GTG Val	CTC Leu	TTT Phe	GAC Asp	GAG Glu	CTT Leu	AGG Arg	CTT Leu	CCC Pro	GCC Ala	TTG Leu	GGG Gly	AAG Lys	ACG Thr	CAA Gln	816
			260					265					270			
AAG Lys	ACG Thr	GGC Gly	AAG Lys	CGC Arg	TCC Ser	ACC Thr	AGC Thr	GCC Ala	GCG Ala	GTG Val	CTG Leu	GAG Glu	GCC Ala	CTA Leu	CGG Arg	864
		275					280					285				

GAG GCC CAC CCC ATC GTG GAG AAG ATC CTC CAG CAC CGG GAG CTC ACC Glu Ala His Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr 290 295 300	912
AAG CTC AAG AAC ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CCG Lys Leu Lys Asn Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro 305 310 315 320	960
AGG ACG GGC CGC CTC CAC ACC CGC TTC AAC ACG GCC ACG GCC ACG Arg Thr Gly Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr 325 330 335	1008
GGG AGG CTT AGT AGC TCC GAC CCC AAC CTG CAG AAC ATC CCC GTC CGC Gly Arg Leu Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg 340 345	1056
ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC GTG GCC GCG GGA Thr Pro Leu Gly Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly 355 360 365	1104
TGG GCG TTG GTG GCC CTG GAC TAT AGC CAG ATA GAG CTC CGC GTC CTC Trp Ala Leu Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu 370 375 380	1152
GCC CAC CTC TCC GGG GAC GAG AAC CTG ATC AGG GTC TTC CAG GAG GGG Ala His Leu Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly 385 390 400	1200
AAG GAC ATC CAC ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG Lys Asp Ile His Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro 405 410 415	1248
GAG GCC GTG GAC CCC CTG ATG CGC CGG GCG GCC AAG ACG GTG AAC TTC Glu Ala Val Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe 420 425 430	1296
GGC GTC CTC TAC GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTT GCC Gly Val Leu Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala 435 440 445	1344
ATC CCC TAC GAG GAG GCG GTG GCC TTT ATA GAG CGC TAC TTC CAA AGC Ile Pro Tyr Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser 450 455 460	1392
TTC CCC AAG GTG CGG GCC TGG ATA GAA AAG ACC CTG GAG GAG GGG AGG Phe Pro Lys Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg 465 470 475 480	1440
AAG CGG GGC TAC GTG GAA ACC CTC TTC GGA AGA AGG CGC TAC GTG CCC Lys Arg Gly Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro 485 490 495	1488
GAC CTC AAC GCC CGG GTG AAG AGC GTC AGG GAG GCC GCG GAG CGC ATG Asp Leu Asn Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met 500 505 510	1536
GCC TTC AAC ATG CCC GTC CAG GGC ACC GCC GCC GAC CTC ATG AAG CTC Ala Phe Asn Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu 515 520 525	1584



- 91 -

GCC ATG GTG AAG CTC TTC CCC CGC CTC CGG GAG ATG GGG GCC CGC ATG Ala Met Val Lys Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met 530 535 540	1632
CTC CTC CAG GTC CAC GAC GAG CTC CTC CTG GAG GCC CCC CAA GCG CGG Leu Leu Gln Val His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg 545 550 555 560	1680
GCC GAG GAG GTG GCG GCT TTG GCC AAG GAG GCC ATG GAG AAG GCC TAT Ala Glu Glu Val Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr 565 570 575	1728
CCC CTC GCC GTG CCC CTG GAG GTG GAG GTG GGG ATG GGG GAG GAC TGG Pro Leu Ala Val Pro Leu Glu Val Glu Val Gly Met Gly Glu Asp Trp 580 585 590	1776
CTT TCC GCC AAG GGT TAG Leu Ser Ala Lys Gly 595	1794

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 597 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Glu Asp Leu Arg Leu Ser Leu Glu Leu Ser Arg Val Arg Thr Asp 1 5 10 15
Leu Pro Leu Glu Val Asp Leu Ala Gln Gly Arg Glu Pro Asp Arg Glu 20 25 30
Gly Leu Arg Ala Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His 35 40 45
Glu Phe Gly Leu Leu Glu Ala Pro Ala Pro Leu Glu Glu Ala Pro Trp 50 55 60
Pro Pro Pro Glu Gly Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu 65 70 75 80
Pro Met Trp Ala Glu Leu Lys Ala Leu Ala Ala Cys Arg Asp Gly Arg 85 90 95
Val His Arg Ala Ala Asp Pro Leu Ala Gly Leu Lys Asp Leu Lys Glu 100 105 110
Val Arg Gly Leu Leu Ala Lys Asp Leu Ala Val Leu Ala Ser Arg Glu 115 120 125
Gly Leu Asp Leu Val Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu 130 135 140
Leu Asp Pro Ser Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly 145 150 155 160

- 92 -

Gly Glu Trp Thr Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg  
 165 170 175  
 Leu His Arg Asn Leu Leu Lys Arg Leu Glu Gly Glu Glu Lys Leu Leu  
 180 185 190  
 Trp Leu Tyr His Glu Val Glu Lys Pro Leu Ser Arg Val Leu Ala His  
 195 200 205  
 Met Glu Ala Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu  
 210 215 220  
 Ser Leu Glu Leu Ala Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe  
 225 230 235 240  
 Arg Leu Ala Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu  
 245 250 255  
 Arg Val Leu Phe Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln  
 260 265 270  
 Lys Thr Gly Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg  
 275 280 285  
 Glu Ala His Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr  
 290 295 300  
 Lys Leu Lys Asn Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro  
 305 310 315 320  
 Arg Thr Gly Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr  
 325 330 335  
 Gly Arg Leu Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg  
 340 345 350  
 Thr Pro Leu Gly Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly  
 355 360 365  
 Trp Ala Leu Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu  
 370 375 380  
 Ala His Leu Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly  
 385 390 395 400  
 Lys Asp Ile His Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro  
 405 410 415  
 Glu Ala Val Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe  
 420 425 430  
 Gly Val Leu Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala  
 435 440 445  
 Ile Pro Tyr Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser  
 450 455 460  
 Phe Pro Lys Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg  
 465 470 475 480  
 Lys Arg Gly Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro  
 485 490 495

- 93 -

[illegible]

## (2) INFORMATION FOR SEQ ID NO:5:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 560 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	Leu
1				5					10					15	
Glu	Ala	Pro	Ala	Pro	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	Gly
			20					25					30		
Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Pro	Glu	Pro	Met	Trp	Ala	Glu
		35					40					45			
Leu	Lys	Ala	Leu	Ala	Ala	Cys	Arg	Asp	Gly	Arg	Val	His	Arg	Ala	Ala
	50					55					60				
Asp	Pro	Leu	Ala	Gly	Leu	Lys	Asp	Leu	Lys	Glu	Val	Arg	Gly	Leu	Leu
65					70					75					80
Ala	Lys	Asp	Leu	Ala	Val	Leu	Ala	Ser	Arg	Glu	Gly	Leu	Asp	Leu	Val
				85					90					95	
Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	Asn
			100					105					110		
Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	Glu
		115					120					125			
Asp	Ala	Ala	His	Arg	Ala	Leu	Leu	Ser	Glu	Arg	Leu	His	Arg	Asn	Leu
	130					135					140				

- 94 -

Leu Lys Arg Leu Glu Gly Glu Lys Leu Leu Trp Leu Tyr His Glu  
 145 150 155 160  
 Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala Thr Gly  
 165 170 175  
 Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Leu Ala  
 180 185 190  
 Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala Gly His  
 195 200 205  
 Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp  
 210 215 220  
 Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly Lys Arg  
 225 230 235 240  
 Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile  
 245 250 255  
 Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys Asn Thr  
 260 265 270  
 Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly Arg Leu  
 275 280 285  
 His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser  
 290 295 300  
 Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln  
 305 310 315 320  
 Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu Val Ala  
 325 330 335  
 Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly  
 340 345 350  
 Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile His Thr  
 355 360 365  
 Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp Pro  
 370 375 380  
 Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu Tyr Gly  
 385 390 395 400  
 Met .Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu  
 405 410 415  
 Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg  
 420 425 430  
 Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly Tyr Val  
 435 440 445  
 Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala Arg  
 450 455 460  
 Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro  
 465 470 475 480

- 95 -

Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys	Leu
				485					490					495	
Phe	Pro	Arg	Leu	Arg	Glu	Met	Gly	Ala	Arg	Met	Leu	Leu	Gln	Val	His
			500					505					510		
Asp	Glu	Leu	Leu	Glu	Ala	Pro	Gln	Ala	Arg	Ala	Glu	Glu	Val	Ala	
			515				520				525				
Ala	Leu	Ala	Lys	Glu	Ala	Met	Glu	Lys	Ala	Tyr	Pro	Leu	Ala	Val	Pro
		530			535					540					
Leu	Glu	Val	Glu	Val	Gly	Met	Gly	Glu	Asp	Trp	Leu	Ser	Ala	Lys	Gly
545					550					555					560

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 24 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCCAGGGTT TTCCAGTCA CGAC

24

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 24 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGCGGATAAC AATTTCACAC AGGA

24

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

- 96 -

CTAAGTAGCT CCGATCCCAA C

21

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATCACTCCTT GCGGAGAGC CAGTC

25

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATTAGCACA TATGCGGATG CTCC

26

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CTTCCAGCT CCGACCCCAA C

21

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 97 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCTACTCCTT GCGGAGAGC CAGTC

25

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGGATGTCCC TCCCCTCTG AAAGA

25

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCCTTTCCCG GAAGCTTTCC CAGGTGCA

28

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TGCACCTGGG AAAGCTTCCG GGAAAGGG

28

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 98 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCTGCAGTAC CGGGAGCTCA CCAAGCTCAA

30

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTGAGCTTGG TGAGCTCCCG GTACTGCAGG

30

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TGGACTATAG CCAGATAGAG CT

22

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AAGCGAAGAC CTCCTCCTCG A

21

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear



- 99 -

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGTTCGGCAG CCTCCTCCAC GA

22

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TCCAAGGAAA GCCTGAGGTC TT

22

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

AAGCTCGCCA TGGTGAAGCT CTT

23

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCGGAGACGA GTTGGTAGAG GT

22

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single

- 100 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ACCTCTACCA ACTCGTCTCC GA

22

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AGAGGACGAA GCCCACGAA

19

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AGGAGGTAGG CGAGGAGCAT

20

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATGCTCCTCG CCTACCTCCT

20

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:

- 101 -

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TCGAGGAGGA GGTCTTCGCT T

21

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AGCTCTATCT GGCTATAGTC CA

22

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATAGGCTCTC CCAGGAGCTT

20

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AAGAGCTTCA CCATGGCGAG CTT

23

- 102 -

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TTCCCTGGA GCGTTTCTG A

21

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AAAGACCACG AAGACGGCCT T

21

## (2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AAGGCCGTCT TCGTGGTCTT T

21

## (2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

- 103 -

AAGGAGTGGG GAAGCCTGGA A

21

## (2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TTCCAGGCTT CCCCACTCCT T

21

## (2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

TTCTTCCGAA GAGGGTTTCC A

21

## (2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GCGTCCAGGA GCGCCCTGAG GA

22

## (2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 104 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CCTCAGGGCG CTCCTGGACG CCA

23

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TTCGTCCTCT CCGCCCCGA

20

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CCAACCTGCA GAACATCCCC GT

22

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GGTGTGGATG TCCTTCCCCT

20

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 105 -

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CCCTGCGGTT TAGAGGAAGT TCAAG

25

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CTTGAAGTTC CTCTAAACGG CAGGG

25

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

ACCCGGCCTT TGGGTTCAAA GA

22

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TCTTTGAACC CAAAGGCCGG GT

22

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

- 106 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TTCCCGTGCT CCTCCGCTC

20

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CTCGCCTTC TCGTCCCTT

20

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GCTTCGGCT CGTATGTTGT GTG

23

(2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 70 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GGAAAGCCTG AGGTCTTCCA TAGCTGTTTC CTGTGTGAAA TTGTTATCCG CTCACAATC  
CACACAACAT

60

70

(2) INFORMATION FOR SEQ ID NO:51:



- 107 -

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 81 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

ACCCGGCCTT TGGGTTCAAA GAGCGGAACG ATCGCCTCCA TAGCTGTTTC CTGTCTGAAA	60
TTGTTATCCG CTCACAATTC C	81

What is Claimed is:

1. A purified DNA in accordance with SEQ ID NO: 3, said DNA encoding for a member selected from the group consisting of a polypeptide in accordance with SEQ ID NO: 5, and fragments thereof having polymerase activity.
2. The DNA of claim 1 consisting of nucleotides 112 to 1791 of SEQ ID NO: 3.
3. The DNA of claim 1 consisting of nucleotides 1 to 1791 of SEQ. ID NO: 3.
4. A vector wherein the DNA of SEQ. ID NO: 3 is operably linked to a promoter.
5. Plasmid p21EHcM1.1, having ATCC Accession No. 69632.
6. A host cell transformed with a DNA selected from the group consisting of the DNAs of claims 1, 2, and 3.
7. The host cell of claim 6, wherein said host cell is capable of expressing a thermostable polypeptide encoded by said DNA, said polypeptide having DNA polymerase activity.
8. The host cell of claim 7, wherein said host cell is a prokaryotic cell.
9. The host cell of claim 8, wherein said host cell is an *E. coli* cell.

10. An expression vector operably linked to nucleotides 112 to 1791 of SEQ ID NO: 3, said nucleotides encoding a polypeptide having thermostable DNA polymerase activity.

11. The expression vector of claim 10 having at least one  
5 insert consisting essentially of nucleotides 112 to 1791 of SEQ ID NO:3.

12. A purified fragment of *Thermus flavus* DNA polymerase I protein in accordance with SEQ ID NO: 5, said fragment having thermostable DNA polymerase activity.

13. A fragment of *Thermus flavus* DNA polymerase I having  
10 thermostable DNA polymerase activity and consisting of amino acids 2 to 560 of SEQ. ID NO: 5.

14. A purified fragment of *Thermus flavus* DNA polymerase I protein encoded by the insert of plasmid p21EHcM1.1, having ATCC Accession no. 69632.

15. The purified fragment of claim 14 wherein the fragment has a DNA polymerase activity between 60,000 U/mg protein and 600,000 U/mg protein.

16. A thermostable polypeptide having DNA polymerase activity, said polypeptide consisting essentially of the amino acid sequence of  
20 SEQ ID NO: 5.

17. A method for purifying a thermostable polypeptide having DNA polymerase activity comprising the steps of:

transforming a host cell with a DNA to create a transformed host cell, said DNA encoding for a thermostable polypeptide having DNA polymerase activity and being selected from the group consisting of the DNAs of claims 1, 2, and 3;

cultivating said transformed host cell under conditions to promote expression of a thermostable polypeptide encoded by said DNA; and

purifying said thermostable polypeptide with a monoclonal antibody that is immunologically cross-reactive with said thermostable polypeptide.

18. The method of claim 17 wherein the host cell is transformed with the DNA of claim 3.

19. The method of claim 17 wherein said immunologically cross-reactive monoclonal antibody has specificity for a *Thermus aquaticus* DNA polymerase.

20. A method of purifying a thermostable polypeptide having DNA polymerase activity comprising the steps of:

- a) expressing said thermostable polypeptide in a host cell, said polypeptide having an amino acid sequence encoded by a DNA selected from the group consisting of the DNAs of claims 1, 2, and 3;
- b) lysing the cell to create a suspension containing said thermostable polypeptide and host cell proteins and cell debris;
- c) contacting a soluble portion of said suspension with an antibody that is immunologically cross-reactive with said thermostable polypeptide and under conditions wherein the antibody binds to said thermostable polypeptide to form an antibody-polypeptide complex;
- d) isolating the antibody-polypeptide complex; and
- e) separating said thermostable polypeptide from said isolated antibody-polypeptide complex to provide a purified thermostable polypeptide.

21. The method of claim 20 further comprising between steps (b) and (c) the steps of:
- heating said suspension to denature host cell proteins; and
  - centrifuging said suspension to remove said cell debris and denatured host cell proteins.

22. The method of claim 20 or 21 wherein said immunologically cross-reactive antibody is a monoclonal antibody.

23. The method of claim 22 wherein said immunologically cross-reactive monoclonal antibody is specific for *Thermus aquaticus* DNA polymerase I.

24. The method of claim 22 wherein the purified thermostable polypeptide has a DNA polymerase activity between 79,500 U/mg protein and 600,000 U/mg protein.

- 5 25. The method of claim 22 wherein the purified thermostable polypeptide has a DNA polymerase activity between 217,600 U/mg protein and 600,000 U/mg protein.

1 / 25

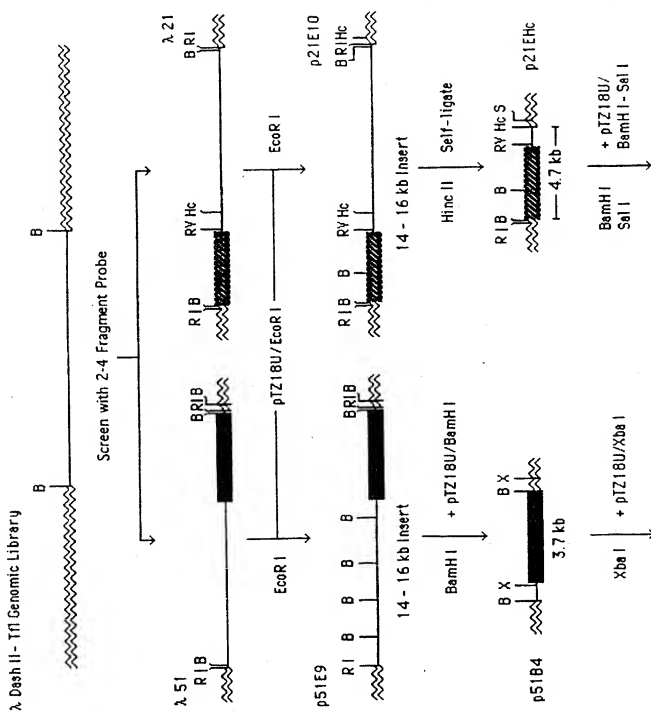


Figure 1A(ii)

2 / 25

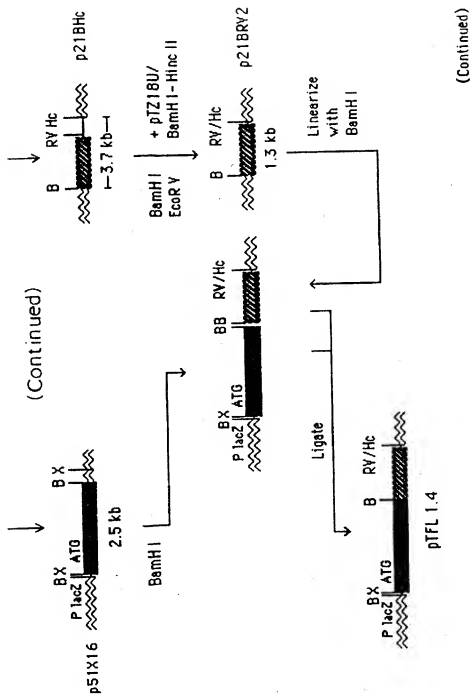


Figure 1A(ii)



3 / 25

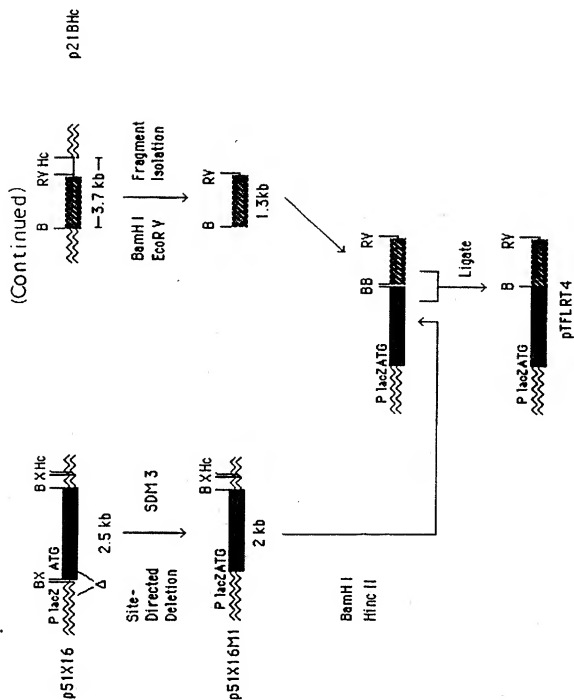


Figure 1A(iii)

4 / 25

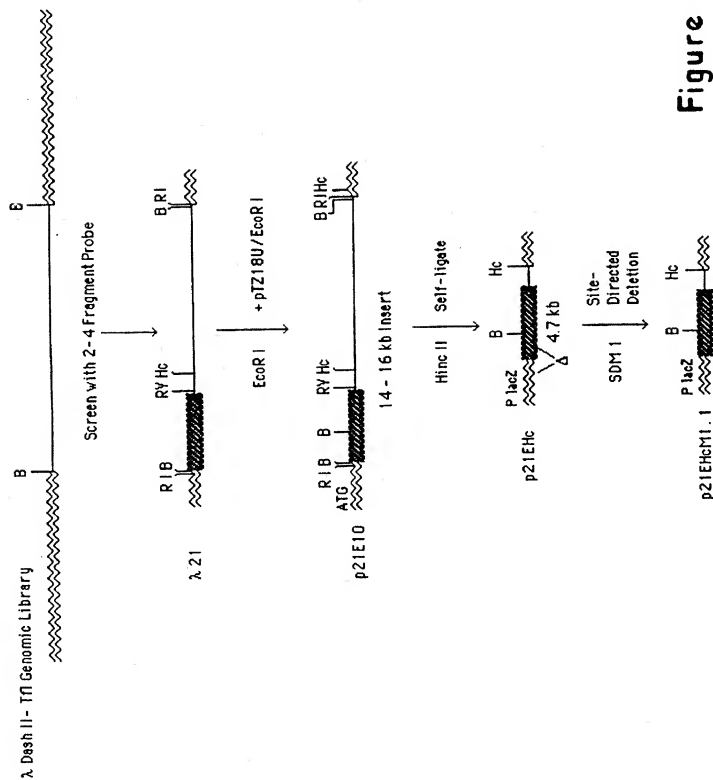


Figure 1B

120 TACTTCCCGGGGTGAAGCTCGGGGCTTGTGCGGCTACGGGGGCTTAAAGCGCGGCCCAAGCTCCCTGTGTGAGCGGGTGGGCTCCCTTC  
 240 TCTGCGCTTCGGCGAGTGGCGGGTCTACGCCCTCTCGAGGCCCGGCTCAAGGCCGAGCAGACCTACACCCCGAGGGCGTGGCTCCCTCTCTCCCTCCCAAGCCCGAGC  
 345 GCGAAGCTTCTCAGCGGGCTCTGGAGCCGACCCGGGACAGGTGGCGCTGGAGTAGATG GAG CGG ATC GTT CCG CTC TTT GAA CCC AAA GCG CGG CTC CTC  
 15 H E A I V P L F E P K G R V L  
 435 CTG GTG GAC GCG GAC CAG CTG GCG TAC CCG ACC CTC ACC AGC CGG GCG GAA CCG GTG CAG GCG GTG CAG GCG GTG TAC  
 45 L V D G H L A Y R T F A L K G L I T S R G E P V Q A V Y  
 525 GCG TTC GCG AAG AGC CTC CTC AAG GCG CTC AAG GCG GAG GCG TAC AAG GCG CTC TTC GTG TTC GAC GCG AAC GCG CCC TCC TTC CTC  
 75 G F A K S L L K A L K E D G Y K A V F V F D A K A P S F R  
 615 CAG GAG GCG TAC GAG GCG TAC AAG GCG GCG ACC CCG GAG GAC TTC CCG GAG CAG CTC CCG CTC ATC AAG GAG CAG GTG GTG GAC  
 105 H E A Y E A Y K A G R A P T P E D F P R Q L I K E L V D  
 705 CTC CTC GCG TTT ACC CCG CTC GAG GTC CCG GCG TAC GAG GCG GAC GTC CTC GCG ACC CTC GCG AAG AAG CCG GAA AAG GAG GCG TAC  
 135 L L G F T R L E V P G Y E A D V L A T L A K K A E K E G Y  
 795 GAG GTG CCG ATC CTC ACC CCG GAC CCG GAC CTC TAC CAA CTC CTC TCC GAG CCG GTG CTC CTC CAC CCG GAG GCG CAC CTC ATC ACC  
 165 E V D L T A D R D L Y Q L V S D R V V V L H P E G H L I T  
 195 CCG GAG TGG CTT TGG GAG AAG TAC GCG CTC AAG CCG GAG CAG TGG GTG GAG TTC CCG GCG CTC GTG GCG GAG CCG TCC GAG AAC CTC CCG  
 P E W L W E K Y G L K P E Q W V D F R A L V G D P S D N L P  
 975 GCG CTC AAG GCG ATC GCG GAG AAG ACC CCG CTC AAG CTC CTC AAG GAG TGG GGA AGC CTC GAA AAC CTC CTC AAG AAC CTC GAG CCG GTA  
 225 G V K G I G E K T A L K L L K E W G S L E N L L K N L D R V  
 885 AAG CCA GAA AAC GTC CCG GAG AAG AAG CCG CAG CTC GAA CCG CTC AGG CTT TCC TTG GAG CTC TCC CCG GTG CCG ACC GAG CTC CCG  
 255 K P E N V R E K A I K A H L E D L R L S L E L S R V R T D L P  
 1155 CTG GAG GTG GAC CTC GCG CAG GCG GCG GCG CCG GTT AGG CCG TTC CTG GAG AGG CTG GAG TTC GCG ACC CTC CTC CAC  
 245 L E V D L A Q G R E P D R E G L R A F L E R L E F G S L L H  
 1245 GAG TTC GCG CTC CTC GAG GCG CCG CTC GAG GAG GCG CCG TGG CCG CCG GCG GCG GCG TTC GTG GCG TTC CTC CTC TCC CCG  
 315 E F G L L E A P A P L E E A P P P E G A F V G F V L S R  
 1335 CCC GAG CCC ATG TGG CCG GAG CTT AAA CCG CTC CCG CTC TCG AGG CAG CCG GCG CTC GAG CCG CCA GAG CCG TTC CCG CCG CTA AAG  
 345 P E P M W A E L K A L A A C R G G R V H X A A D P L A G L K

(Continued)

Figure 2 (i)

(Continued)

GAC CTC AAG GTC CTC GGT CTC CTC AAG GAC CTC CCC TTG CCC TCG AGG GAG GGG CTA GAC CTC CTC CCC GGG GAC GAC CCC  
 D L K E V R G L L A K D L A V L A S R E G L D L V P G D P  
 1425  
 375  
 ATG CTC CTC GGC TAC CTC GAC CCC TCC AAC ACC ACC CCG GAG GGG GTG CCG CCG TAC CCG GGG GAG TGG AGG GAC GAC GGC CCC  
 H L L A Y L L D P S N T T P E G V A R R Y G G E W T E D A A  
 1515  
 405  
 CAC CCG CCC CTC CTC TCG GAG AGG CTC CAT CCG AAC CTC CTT AAG CCC CTC GAG GGG GAG AAG CTC CTT TGG CTC TAC CAC GAG GTG  
 H R A L L S E R L H R N L L K R L E G E K L L W L Y H E V  
 1605  
 435  
 GAA AAG CCC CTC TCC CCG GTC CCG CAC ATG CAC GGC ACC CCG GTA CCG CTG GAC GTG GGC TAC CTG CAG GGC CTT TCC CTC GAG CTT  
 E K P L S R V L A H M E A T G V R L D V A Y L Q A L S L E L  
 1695  
 465  
 CCG GAG ATC CCG CTC GAG GAG GAG GTC TTC CCG GGC CAC CCC TTC AAC CTC AAC TCG CCG GAC CAG CTG GAA AGG GTG  
 A E E I R R L E E V F R L A G H P F N L N S R D Q L E R V  
 1785  
 495  
 CTC TTT GAG GAG CTT AGG CTT CCC GGC TTC GGC AAG ACC CAA AAG AGC GGC AAC CCG TCC ACC AGC GGC GGC GTG CTC GAG CCC CTA CCG  
 L F D E L R L P A L G K T Q K T G K R S T S A A V L E A L R  
 1875  
 525  
 GAG CCC CAC CCG ATC CTC CAG AAC ATC CTC CAG CAG CCG GAG CTC ACC AAG CTC AAG AAC ACC TAC GTG CAC CCC CTC CCA AGC CTC CTC  
 E A H P I V E K I L Q H R E L T K L K N T Y V D P L P S L V  
 1965  
 555  
 CAC CCG AGG AGC GGC CTC CAC ACC CCG TTC AAC CAG AGC GGC ACC GGC ACC GGC AGG CTT AGT AGC TCC CAC CCC AAC CTC GAG AAC  
 H P R T G R L H T R F H Q T A T A T I G R L S S S D P N L Q H  
 2055  
 585  
 ATC CCC GTC CCG ACC CCG CAG AGG ATC CCG CCG CTC CTC GGC GAG CCG GGA TCG GCG TTC CTC GGC CTC GAG TAT AGC CAG  
 I P V R T P L G Q R I R R A F V A E A G W A L V A L D Y S Q  
 2145  
 615  
 ATA GAG CTC CCG CTC CCG CAC CTC TCC CCG GAG CAG AAC CTG ATC AGG GTC TTC CAG CAG CCG AAG CAC ATC CAC ACC CAG ACC GCA  
 I E L R V L A H L S G D E H L I R V F Q E G C D I H T Q T A  
 2235  
 645  
 AGC TGG ATG GTC GGC CTC CCG CCG GAG CCG CTC ATG CCG CCG GGC AAC GTC AAC TTC GGC CTC TAC GGC ATG  
 S W M F G V C P P E A V D P L M R A K T V N F G V L Y G H  
 2325  
 675  
 TCG GCG CAT AGG CTC TCC CAG GAG CTT GCG ATC CCG TAC GAG CCG GTG GGC TTT ATA GAG CCG TAC CAA AGC TTC CCC AAG GTG  
 S A H R L S Q E L A I P Y E A V A F I E R Y F Q S F P K V  
 2415  
 705

(Continued)

Figure 2 (ii)

(Continued)

CGG GCC TCG ATA GAA AAG ACC CTG GAG GGG AGG AAG CGG GGC TAC GTG GAA ACC CTC TTC GGA AGA AGG CCG TAC GTG CCC GAC CTC  
 2595  
 R A W I E K T L E E G R K R G Y V E T L F G R R Y V P D L 735  
  
 AAC CCC CGG GTG AAG AGC GTC AGG GAG GCC CGG GAG CGC ATG GGC TAC AAC ATG CCC GTG CAG GGC ACC GGC GAC CTC ATG AAG CTC  
 2595  
 N A R V X S V R E A A E R H A F N H P V Q G T A A D L H X L 765  
  
 GCC ATG GTG AAG CTC TTC CCC CTC CGG GAG ATG GGG GCC CGG ATG CTC CTC CAG GTC CAC GAG CTC CTC GTG GAG CCC CCC CAA  
 2685  
 A H V X L F P R L R E H G A R M L L Q V H D E L L L E A P Q 795  
  
 CGG CGG GGC GAG GTG GCG GCT TTG GCC AAG GAG GGC ATG GAG AAG GCC TAT CCC CTC GCC GTG CCC CTC GAG GTG GAG GTG GGG ATG  
 2775  
 A R A E E V A A L A K E A H E K A Y P L A V P L E V E V G H 825  
  
 CGG GAG GAC TGG CTT TTC GCC AAG GGT TAGGGGGCCCTGCCGTTTACAGGAGTTCAAGGGGTTTCCCTCAGAAAGCCCTCCAGGGAGCCCTCTGCCGTACAGG  
 2886  
 G E D W L S A K G 835  
  
 AGCCCTTTAGCCCCCAAGGTGCGGGTGAAGGCTTCAGGGCCCTGGGTCTCTTTTAAAGGGGGCGCTTTTCACTCAGAGGCCAGAGGGCGCTTTCCTTTTGAGAGCAAGTCACTTCCT  
 3086  
  
 CGTCCCTTTCCGCCCCAGTAGTACACTCAAAACCCCCCTCGT 3048

Figure 2 (iii)

10 20 30 40 50 60 70 80  
 MBR TFL MEAIVLPFKGRVLLVDGHHLAVRTFFALKGLTTSRGEVPQAVYGFASLLKALKEDGYKAVFVVDKAKAPSRHEAYE  
 TAQ MRGMLPLFEPKGRVLLVDGHHLAVRTFFALKGLTTSRGEVPQAVYGFASLLKALKEDG-DAVIVVDKAKAPSRHEAYG  
 A&V TFL MAMPLPLFEPKGRVLLVDGHHLAVRTFFALKGLTTSRGEVPQAVYGFASLLKALKEDG-DVVVVVVDKAKAPSRHEAYE  
 \*^\*\*\*  
 MBR TFL AYKAGRAPTPEDFPQALIKELVDLLGFRLEVPVGYEADVDVATLAKAEKGEYVRILTDADRLDQYLVSDDVVVLHPE  
 TAQ GYKAGRAPTPEDFPQALIKELVDLLGLLARLEVPGYEADVDVATLAKAEKGEYVRILTDADRLDQYLVSDDVVVLHPE  
 A&V TFL AYKAGRAPTPEDFPQALIKELVDLLGLVLRLEVPGYEADVDVATLAKAEKGEYVRILTDADRLDQYLVSDDVVVLHPE  
 \*  
 MBR TFL GHLLTPENLWEKYLCKPEQWDFRALVGDPSDNLPGVKIGCKTALKLLKMGWSENLKLLNLDVRKPNVREKIKAHLED  
 TAQ GYLITPAMLWEKYLCKPRDQWADYRALTGDSNLPVKIGCKTALKLLKMGWSENLKLLNLDVRKPNVREKIKAHMDD  
 A&V TFL GYLITPAMLYEKYLCKPRDQWADYRALTGDSNLPVKIGCKTALKLLKMGWSENLKLLNLDVRKPNVREKIKAHMDD  
 \*  
 MBR TFL LRLSLSRVRTDLPLEVDLAQGREPDREGRAFLEFGLSLLHEFGLEAPAPLEAEAPWPPPEGAPVGVLSRPEPHW  
 TAQ LKLSMDLAKVVRTDLPLEVDFAKRREPDRELRAPLEFGLSLLHEFGLEAPAPLEAEAPWPPPEGAPVGVLSRPEPHW  
 A&V TFL LALSRLKSQVHTDLPLEVDPGRRRTPNLEGLRAFLEFGLSLLHEFGLEAPAPLEAEAPWPPPEGAPVGVLSRPEPHW  
 ^ ^ \* \* \*  
 MBR TFL AELKALACRDGRVHRAADPLAGLKDLEKVRGLLAKDLAVLASREGDLVPGDDPHLLAYLLDPSNTTPEGVARRYGGEW  
 TAQ ADLLALAAARGGRVHRAPEPTKALRDLEKARGLLAKDLAVLASREGDLVPGDDPHLLAYLLDPSNTTPEGVARRYGGEW  
 A&V TFL AELLALAGAWEGRLHRAQDPLRGLRDLKVGRIKAKDLAVLASREGDLVPGDDPHLLAYLLDPSNTTPEGVARRYGGEW  
 \* ^ ^ \* \* \*  
 MBR TFL TEDAHRAALLSERLHRNLLKRLKEGKLLWLYHVEVEKPLSRVLAHMEATGVRLDVAIQLALSLELAEEIRLREEVEVRLA  
 TAQ TTEAGERAALSERLAFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAIQLALSLELAEEIRLREEVEVRLA  
 A&V TFL TEDAGERALLAERLFTQLERLKGERRLLWLYREVEKPLSRVLAHMEATGVRLDVAIQLALSLELAEEIRLREEVEVRLA  
 \* \* \* ^ ^ ^ \* \*  
 MBR TFL GHFPNLSRQDLERVLFDLRLPALCQTKTKRSTSAVLEALREAHPIVEKILQHRELTKLKNYVDPLPSLVHPRTG  
 TAQ GHFPNLSRQDLERVLFDLRLPALCQTKTKRSTSAVLEALREAHPIVEKILQHRELTKLKNYVDPLPSLVHPRTG  
 A&V TFL GHFPNLSRQDLERVLFDLRLPALCQTKTKRSTSAVLEALREAHPIVDRLQILRELTCLKKNYIDPLPALVHPKRTG  
 \* \* \* \* \*  
 (Continued)

Figure 3 (i)

(Continued)

```

MBR TFL RLHTRFNQTATATGRLSSDPNLQNIPTPLGQIRIRAFVAGGMAVLDYSQIELRVLAHLSGDNELIRVQEGKDI
TAQ   RLHTRFNQTATATGRLSSDPNLQNIPTPLGQIRIRAFIAEEGWLIVLDYSQIELRVLAHLSGDNELIRVQEGKDI
A&V TFL RLHTRFNQTATATGRLSSDPNLQNIPTPLGQIRIRAFVAAEGWLVLDYSQIELRVLAHLSGDNELIRVQEGKDI
      * * * ^ *
MBR TFL HTQTASWHFGVPPPEAVDPLMRRAAKTVNFVLYGSAHRLSQELAIPEEZAFAPIERYFSQSPKVRAMIEKTLIEGRRG
TAQ   HTQTASWHFGVPPPEAVDPLMRRAAKTVNFVLYGSAHRLSQELAIPEEZAFAPIERYFSQSPKVRAMIEKTLIEGRRG
A&V TFL HTQTASWHFGVPPPEAVDPLMRRAAKTVNFVLYGSAHRLSQELAIPEEZAFAPIERYFSQSPKVRAMIEKTLIEGRRG
      * * * ^ *
MBR TFL YVETLFGRRRYVPDNLNARVKSUREAAERMAFNMPVQGTADLAKLAHVKLFPRLREMGARMLLQVHDELLLEAPQARAE
TAQ   YVETLFGRRRYVPDNLNARVKSUREAAERMAFNMPVQGTADLAKLAHVKLFPRLREMGARMLLQVHDELLLEAPQARAE
A&V TFL YVETLFGRRRYVPDNLNARVKSUREAAERMAFNMPVQGTADLAKLAHVKLFPRLREMGARMLLQVHDELLLEAPQARAE
      *
MBR TFL VAALAKEANEKAYPLAVPLEVEVCGEDWLSAKG
TAQ   VARLAKEVMEGVYPLAVPLEVEVCGEDWLSAKE
A&V TFL VAALAKEVMEGVYPLAVPLEVEVCGEDWLSAKE
      ^ * * * *

```

\* = Conservative change    ^ = Non-conservative change

Figure 3 (ii)

10 / 25

10 20 30 40 50 60  
 tacttcggcg ggggtgaagct cggggccggg gggcttggtc gggccctacg ggggggtggcg  
 atgaagccgc cccacttcga gccccggccc cccgaacacg cccggatgcc cccccaccgc  
 70 80 90 100 110 FTFLZ 120  
 gcggaggcct taagcgggcg cccaagggtc ccttggtgga gggggtggg ctcgccctcc  
 cgcctccgca attcgcccgc ggggttcagg ggaaccact cggccacccc gagcggaagg  
 130 140 150 160 170 180  
tcgtgccctt cgcgaggtg ggcgggtct acgccctcct ggaggccgc gccctgaagg  
agcacgggaa gcggtccac cgggccaga tgcgggagga cctccggcg cgggacttc  
 190 200 210 220 230 240  
 ccgaggagac ctacacccg gaggcggtc gcttcgcct cctcctccc aagcccgagc  
 ggctcctctg gatgtggggc ctccgcacg cgaagcggga ggaggagggg ttccgggtcg  
 250 FTFLQ 260 270 280 290 FTFL10 300  
 gggaggttt cttcaggcg ctcctggacg ccaccgggg acaggtggcc ctggagtagc  
 cctttccaa ggagtcctgc gaggacctgc ggtgggccc tgtccacgg gacctcatcg  
 RTFLP 310 320 FTFLX 330 340 350 360  
ATGAGGCGA TCGTTCGCT CTTTGAACCC AAAGGCCGGG TCTCTCTGGT GGACGGCCAC  
TACCTCCGCT AGCAAGGCGA GAAACTTGGG TTTCCGGCCC AGGAGGACCA CCTGCCGGTG  
 RTFLW  
 370 380 390 400 410 420  
 CACCTGGCCT ACCGCACCTT CTTCGCCCTG AAGGGCCTCA CCACGAGCCG GGGCGAACCG  
 GTGGACCGGA TGGCGTGGAA GAAGCGGGAC TTCCCGGAGT GGTGCTCGGC CCCGCTTGGC  
 430 440 450 460 470 480  
 GTGACGGCGG TCTACGGCTT CGCCAAGAGC CTCCTCAAGG CCCTGAAGGA GGACGGGTAC  
 CACGTCCGCC AGATGCCGAA GCGGTTCTCG GAGGAGTTTC GGGACTTCTT CTTGCCCATG  
 RTFLY  
 FTFL 490 500 510 520 530 540  
AAGGCCGTCT TCGTGGTCTT TGACGCCAAG GCCCCTCCT TCCGCCACGA GGCCTACGAG  
TTCCGGCAGA AGCACCAGAA ACTGCGGTTT CGGGGGAGGA AGGCGGTGCT CCGGATGCTC  
 RTFLK  
 550 560 570 580 590 600  
 GCCTACAAGG CGGGGAGGGC CCCGACCCC GAGGACTTCC CCCGGCAGCT CGCCCTCATC  
 CGGATGTTC GCCCCCTCCG GGGCTGGGG CTCTGAAGG GGGCCGTGCA GCGGGAGTAG  
 610 620 630 640 650 660  
 AAGGAGCTGG TGGACCTCCT GGGGTTTACC CGCCTCGAGG TCCCCGGCTA CGAGGGCGAC  
 TTCCTCGACC ACCTGGAGGA CCCCAAATGG GCGGAGCTCC AGGGGCCGAT GCTCCGCTG

(Continued)

Figure 4 (i)



11 / 25

(Continued)

670 680 690 700 710 720  
 GACGTCCTCG CCACCTTGGC CAAGAAGGCG GAAAAGGAGG GGTACGAGGT GCGCATCCTC  
 CTGCAGGAGC GGTGGGACCG GTTCTTCCGC CTTTCTCTCC CCATGCTCCA CGCGTAGGAG  
  
 730 FTFLB 740 750 760 770 780  
 ACCGCCGACC GCGACCTCTA CCAACTCGTC TCCGACCGCG TCGTCGTCTT CCACCCCGAG  
 TGGCGGCTGG CGCTGGAGAT GGTGAGCAG AGGCTGCGCG AGCAGCAGGA GGTGGGGCTC  
 RTFLA  
 790 800 810 820 830 840  
 GGCCACCTCA TCACCCCGGA GTGGCTTGG GAGAAGTACG GCCTCAAGCC GGAGCAGTGG  
 CCGGTGGAGT AGTGGGGCCT CACCGAAACC CTCTTCATGC CGGAGTTCCG CCTCGTCACC  
  
 850 860 870 880 890 900  
 GTGGACTTCC GCGCCCTCGT GGGGGACCCC TCCGACAACC TCCCCGGGGT CAAGGGCATC  
 CACCTGAAGG CCGGGGAGCA CCCCCTGGGG AGGCTGTTGG AGGGGCCCCA GTTCCCGTAG  
  
 910 920 930 FTFLM 940 950 960  
 GGGGAGAAGA CCGCCCTCAA GCTCCTAAG GAGTGGGGAA GCCTGGAAAA CCTCCTCAAG  
 CCCCTCTTCT GCGGGGAGTT CGAGGATTC CTACCCCTTT CCGACCTTTT GGAGGAGTTT  
 RTFLN  
 970 980 990 1000 1010 1020  
 AACCTGGACC GGGTAAAGCC AGAAAACGTC CGGGAGAAGA TCAAGGCCCA CCTGGAAGAC  
 TTGGACCTGG CCCATTTCGG TCTTTTGAG GCCCTCTTCT AGTTCCGGGT GGAGCTTCTG  
 1030 FTFL15 1040 1050 1060 1070 1080  
 CTCAGGCTTT CTTTGGAGCT CTCCCGGGTG CGCACCAGCC TCCCCCTGGA GGTGGACCTC  
 GAGTCCGAAA GGAACCTCGA GAGGGCCAC GCGTGGCTGG AGGGGGACCT CCACCTGGAG  
 TFLE1 RTFL16  
 1090 1100 1110 1120 1130 TFLE1 1140  
 GCCCAGGGGC GGGAGCCCGA CCGGGAGGGG CTTAGGGGCT TCCTGGAGAG GCTGGAGTTC  
 CGGGTCCCCG CCTCGGGCT GGCCTCCTCC GAATCCCGGA AGGACCTCTC CGACCTCAAG  
  
 1150 1160 1170 1180 1190 1200  
 GGCAGCCTCC TCCACGAGTT CGGCCTCCTG GAGGCCCCCG CCCCCCTGGA GGAGGGCCCC  
 CCGTCGGAGG AGGTGCTCAA GCCGGAGGAC CTCCGGGGGC GGGGGGACCT CCTCCGGGGG  
  
 1210 1220 1230 FTFLR 1240 1250 1260  
 TGGCCCCCGC CGGAAGGGGC CTTGCTGGGC TTTGTCCTCT CCGGCCCCGA GCCCATGTGG  
 ACCGGGGGCG GCCTTCCCCG GAAGCACCCG AAGCAGGAGA GGGCGGGGCT CGGGTACACC  
 RTFLC  
 1270 1280 1290 1300 1310 1320  
 GCGGAGCTTA AAGCCCTGGC CGCCTGCAGG GACGGCCGGG TGACCCGGGC AGCAGACCCC  
 CGCCTCGAAT TTCGGGACCG GCGGACGTCC CTGCCGGCCC ACGTGCCCCG TCCTCTGGGG

(Continued)

Figure 4 (ii)

12 / 25

(Continued)

1330	1340	1350	1360	1370	1380
TTGGCGGGG	TAAAGGACCT	CAAGGAGGTC	CGGGGTCTCC	TCGCCAAGGA	CCTCGCCGTC
AACCGCCCCG	ATTTCTTGGA	GTTCTCCAG	GCCCCAGAGG	AGCGGTTCTT	GGAGCGGCAG
1390	1400	1410	1420	1430	1440
TTGGCCTCGA	GGGAGGGGCT	AGACCTCGTG	CCCCGGGACG	ACCCATGCT	CCTCGCCTAC
AACCGGAGCT	CCCTCCCCGA	TCTGGAGCAC	GGGCCCCCTGC	TGGGATACGA	GGAGCGGATG
				RTFLD	
1450	1460	1470	1480	1490	1500
CTCCTGGACC	CCTCCAACAC	CACCCCGGAG	GGGGTGGCGC	GGCGTACGG	GGGGGAGTGG
GAGGACCTGG	GGAGGTTGTG	GTGGGGGGCTC	CCCCACCGCG	CCGCGATGCC	CCCCCTCACC
1510	1520	1530	1540	1550	1560
ACGGAGGACG	CGGCCACCG	GGCCCTCCTC	TCGGAGAGGC	TCCATCGGAA	CCTCCTTAAG
TGCCTCTGCG	GGCGGGTGCG	CCGGGAGGAG	AGCCTCTCCG	AGGTAGCCTT	GGAGGAATTC
1570	1580	1590	1600	1610	1620
CGCCTCGAGG	GGGAGGAGAA	GCTCCTTTGG	CTCTACCACG	AGGTGGAAAT	GGCCCTCTCC
GCGGAGCTCC	CCCTCCTCTT	CGAGGAAACC	GAGATGGTGC	TCCACCTTTT	CGGGAGAGAG
1630	1640	1650	1660	1670	1680
CGGGTCTTGG	CCACATGGA	GGCCACCGGG	GTACGGCTGG	ACGTGGCCTA	CCTCGAGGCC
GCCCAGGACC	GGGTGTACTT	CCGGTGGCCC	CATGCCGACC	TGCACCGGAT	GGACGTCCGG
1690	1700	1710	1720	1730	1740
CTTTCCCTGG	AGCTTGCGGA	GGAGATCCGC	CGCCTCGAGG	AGGAGGTCCT	CCGCTTGGCG
GAAAGGGACC	TGGAACGCCT	CCTCTAGGCG	GCGGAGCTCC	TCCTCCAGAA	GGCGAACCGC
			RTFL18A		
1750	1760	1770	1780	1790	1800
GGCCACCCCT	TCAACCTCAA	CTCCCCGGAC	CAGCTGGAAA	GGGTGCTCTT	TGACGAGCTT
CCGGTGGGGA	AGTTGGAATT	GAGGGCCCTG	GTCGACCTTT	CCCACGAGAA	ACTGCTCGAA
1810	1820	1830	1840	1850	1860
AGGCTTCCCG	CCTTGGGGAA	GACGCAAAAG	ACGGGCAAGC	GCTCCACCAAG	CGCCGCGGTT
TCCGAAGGGC	GGAACCCCTT	CTGCGTTTTT	TGCCCGTTTC	CGAGGTGGTC	GCGGGCCGAC
1870	1880	1890	1900	FTFL17	1910
CTGGAGGCCC	TACGGGAGGC	CCACCCCATC	GTGGAGAAGA	TCCTCCAGCA	CCGGGAGCTC
GACCTCCGGG	ATGCCCTCCG	GGTGGGGTAG	CACCTCTTCT	AGGAGGTCGT	GGCCCTCCAG
				RTFL18	
1930	1940	1950	1960	1970	1980
ACCAAGCTCA	AGAACACCTA	CGTGGACCCC	CTCCCAAGCC	TCGTCCACCC	GAGGACGGGC
TGGTTCGAGT	CTTGTGGAT	GCACCTGGGG	GAGGGTTCCG	AGCAGGTGGG	CTCTCGCCCG
				FTFL2	
1990	2000	2010	2020	2030	FTFL11
CGCCTCCACA	CCCGCTTCAA	CCAGACGGCC	ACGGCCACGG	GGAGGCTTAG	TAGCTCCGAC
GCGGAGGTGT	GGGCGAAGTT	GGTCTGCCGG	TGCCGGTGCC	CCTCCGAATC	ATCGAGGCTG

Figure 4(iii)

(Continued)

(Continued)

13/25

2050	FTFLS	2060	2070	2080	2090	2100
CCCAACTGCG	AGAACATCCC	CGTCCGCACC	CCCTTGGGCC	AGAGGATCCG	CCGGGCGCTTC	
GGGTGGACG	TCTTGTAGGG	GCAGGCGTGG	GGGAACCCGG	TCTCCTAGGC	GGCCCGGAAG	

2110	2120	2130	FTFL17A2140	2150	2160
GTGGCCGAGG	CGGGATGGGC	GTTGGTGGCC	GTGGACTATA	GCCAGATAGA	GCTCCGCGTC
CACCGGTCTC	GCCCTACCCG	CAACCACCCG	ACCTGATAT	CGGTCTATCT	CGAGGCGCAG

RTFLG

2170	2180	2190	2200	2210	2220
CTCGCCACCC	TCTCCGGGGA	CGAGAACCCTG	ATCAGGGTCT	TCCAGGAGGG	GAAGGACATC
GAGCGGTGG	AGAGGCCCT	GCTCTTGAC	TAGTCCAGA	AGTCTCTCCC	CTTCTGTAG

RTFL13      RTFLT

2230	2240	2250	2260	2270	2280
CACACCCAGA	CCGCAAGCTG	GATGTTCCGG	GTCCCCCCGG	AGGCCGTGGA	CCCCCTGATG
GTGTGGTCT	GGCGTTGAC	CTACAAGCCG	CAGGGGGGCC	TCCGGCACCT	GGGGGACTAC

2290	2300	2310	2320	2330	FTFLH	2340
CGCGCGCGCG	CCAAGACGGT	GAACCTTCGG	GTCCTCTACG	GCATGTCCCG	CQATAGGCTC	
CGGCGCCGCC	GGTTCGCCA	CTTGAAGCCG	CAGGAGATGC	CGTACAGCGC	GGTATCCGAG	

2350	2360	2370	2380	2390	2400
TCCAGGAGC	TGGCCATCCC	CTACGAGGAG	GCGGTGGCCT	TTATAGAGCG	CTACTTCCAA
AGGGTCTCG	AACGGTAGGG	GATGCTCCTC	CGCCACCCGA	AAATATCTCG	GATGAAGGTT

2410	2420	2430	2440	2450	2460
AGCTTCCCA	AGGTGCGGGC	CTGGATAGAA	AAGACCCTGG	AGGAGGGGAG	GAAGCGGGGC
TGGAAGGGGT	TCCACGCCCG	GACCTATCTT	TTCTGGGACC	TCCTCCCTCT	CTTCGCCCGC

2470	2480	2490	2500	2510	2520
TACGTGAAAA	CCCTCTTCGG	AAGAAGGCGC	TACGTGCCCG	ACCTCAACGC	CCGGGTGAAG
ATGCACTTT	GGGAGAAGCC	TTCTTCCGCG	ATGCACGGGC	TGGAGTTGCG	GGCCCACTTC

RTFL0

2530	2540	2550	2560	2570	2580
AGCGTCAGGG	AGGCCGCGGA	GCGCATGGCC	TTCAACATGC	CCGTCCAGGG	CACCGCCGCC
TCCGAGTCCC	TCCGGCGCCT	CGCGTACCGG	AAGTTGTACG	GGCAGGTCCC	GTGGCGGCGG

2590	FTFLSF12600	2610	2620	2630	2640
GACCTCATGA	AGCTCGCCAT	GGTGAAGCTC	TTCCCCCGCC	TCCGGGAGAT	GGGGGCCCGC
CTGGAGTAQT	TCCAGCGGTA	CCACTTCGAG	AAGGGGGCGG	AGGCCCTCTA	CCCCCGGGCG

RTFLI

2650	2660	2670	2680	2690	2700
ATGCTCTCC	AGGTCCACGA	CGAGCTCCTC	CTGGAGGGCC	CCCAAGCGCG	GGCCGAGTGG
TACGAGGAGG	TCCAGGTGCT	GCTCGAGGAG	GACCTCCGGG	GGGTTCGCGC	CCGGCTCCTC

(Continued)

Figure 4 (iv)

14 / 25

(Continued)

2710	2720	2730	2740	2750	2760
GTGGCGGCTT	TGGCCAAGGA	GGCCATGGAG	AAGGCCTATC	CCCTCGCCGT	GCCCCCTGGAG
CACCGCCGAA	ACCGGTTCT	CCGTACCTC	TTCCGGATAG	GGGAGCGGCA	CGGGGACCTC
2770	2780	2790	2800	2810	RTFLU 2820
GTGGAGGTGG	GGATGGGGGA	GGACTGGCTT	TCCGCCAAGG	GTTAGggggg	ccttgccgtt
CACCTCCACC	CCTACCCCT	CTGACCGAA	AGGCGGTTCC	CAATCcccc	gggacggcaa
2830	2840	RTFL4/RTFL12	2850	2860	2870
tagaggaagt	tcaagggtt	gtccctcaga	aacgcctcca	gggaaacgcc	ctctgcggct
atctcctcca	agttccccaa	cagggagtct	ttgcggaggt	ccccttgccg	gagacgccga
2890	2900	RTFLJ	2910	2920	2930
accaggaggc	ctttagcccc	aaaggtgcgg	gtgaaggctt	ccaggccctg	ggttctttta
tgtcctccg	gaaatcgggg	tttccacgcc	cacttcggaa	ggtccgggac	ccaagaaat
2950	2960	2970	2980	2990	3000
aagggggcgc	ttttgacctc	gagggccagg	agggcctttc	ccttttgaag	gacaaagtca
ttcccccgcg	aaaactggag	ctccgggtcc	tccgcgaaag	ggaaaacttc	ctgtttcagt
3010	3020	3030	3040	3050	3060
cttcctggtc	cctttccccc	cagtagtaca	cctcaaaccc	cccctgggt..	.....
gaaggaccag	ggaagggcgc	gtcatcatgt	ggagtttggg	ggggacca..	.....

Figure 4(v)

15 / 25

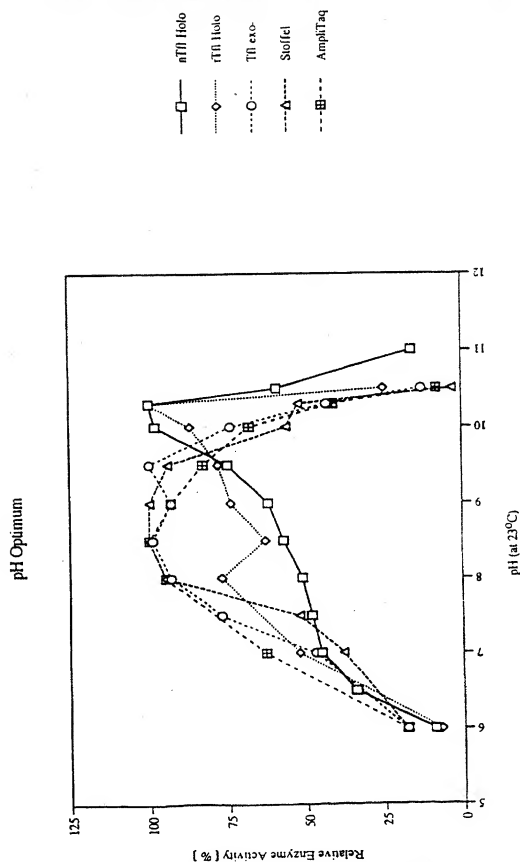


Figure 5

16 / 25

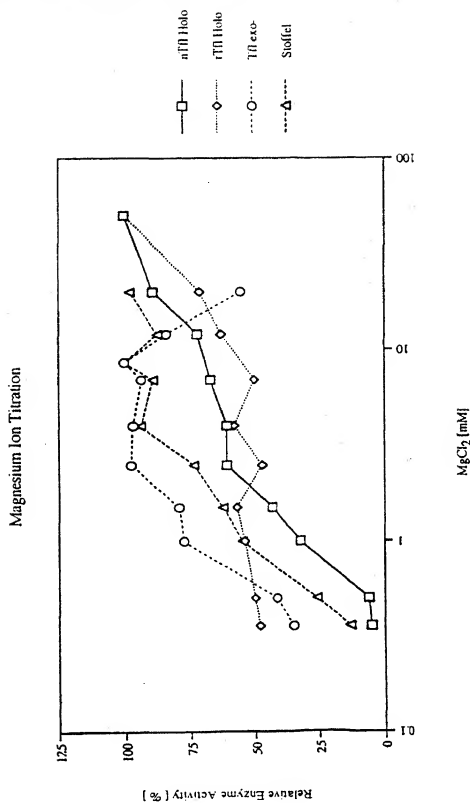


Figure 6A

17 / 25

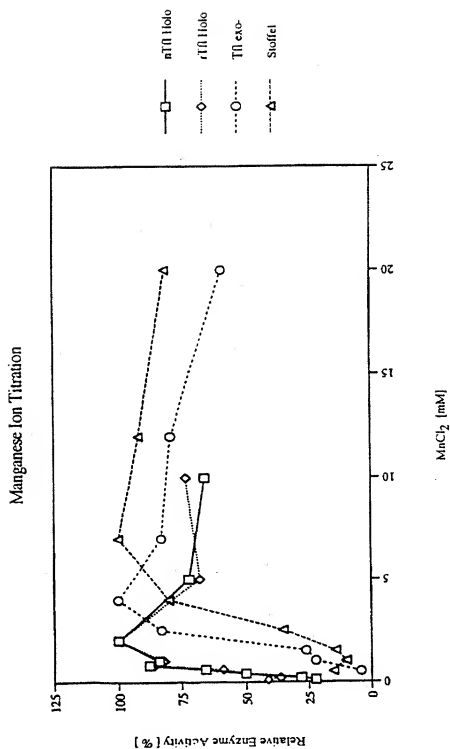
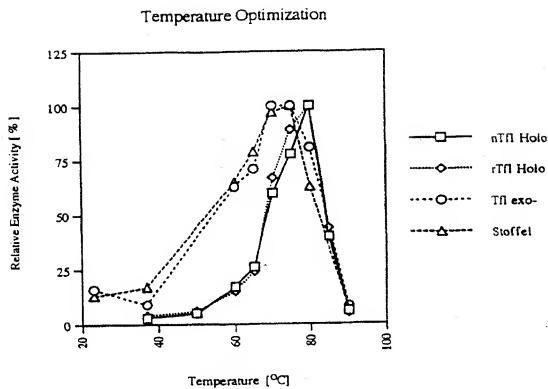


Figure 6B

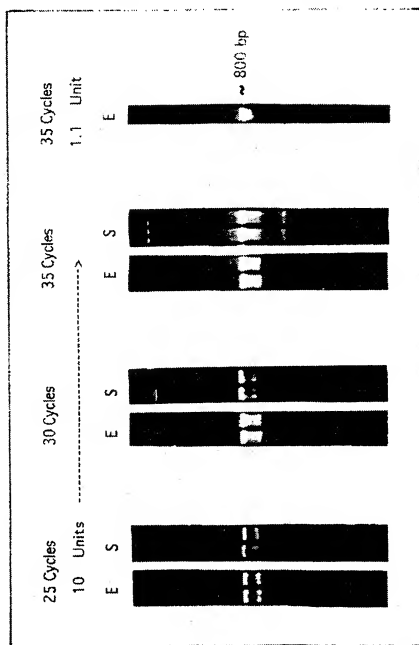
18 / 25

**Figure 7A**



19 / 25

Figure 7B



20 / 25

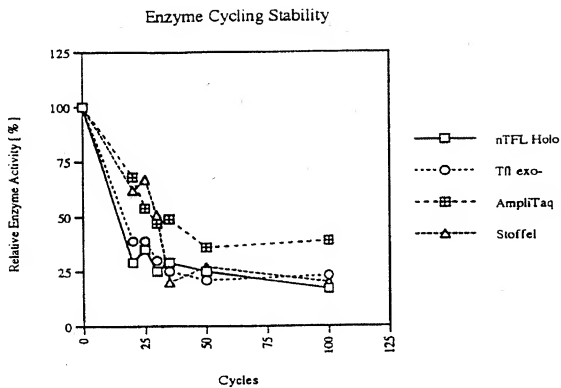


Figure 8

21 / 25

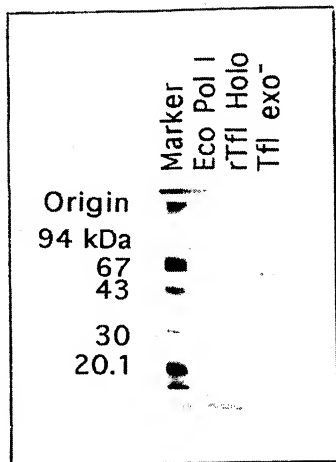
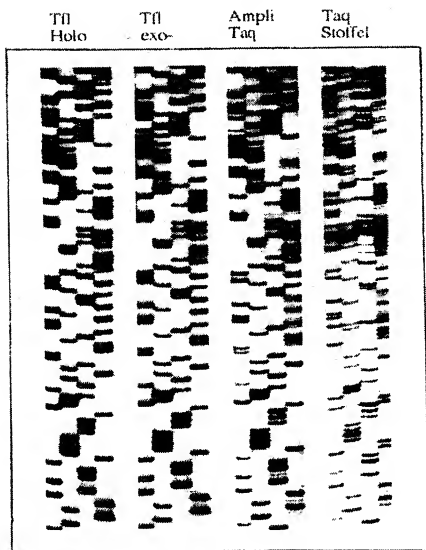


FIGURE 9

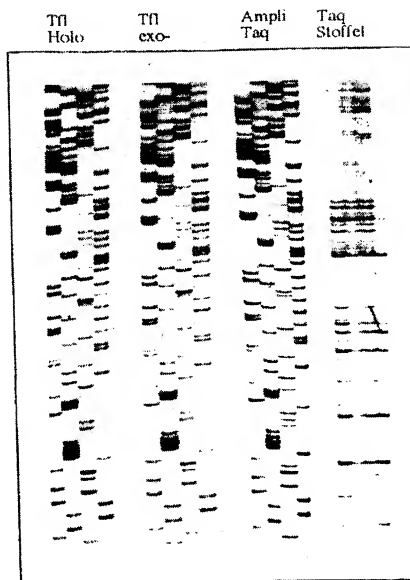
22 / 25

Figure 10A



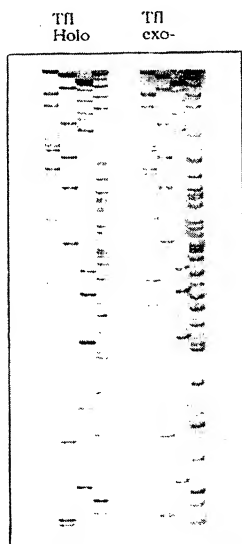
23 / 25

Figure 10B



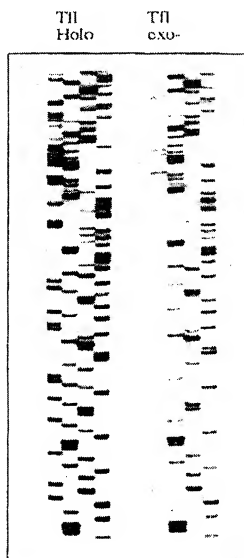
24 / 25

Figure 10C



25 / 25

Figure 10D



# INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

<b>A. The indications made below relate to the microorganism referred to in the description</b> on page <u>80</u> , line <u>1-10</u> (and elsewhere)	
<b>B. IDENTIFICATION OF DEPOSIT</b> <span style="float: right;">Further deposits are identified on an additional sheet <input type="checkbox"/></span>	
Name of depositary institution American Type Culture Collection (ATCC)	
Address of depositary institution (including postal code and country) 12301 Parklawn Dr. Rockville, MD 20852 USA	
Date of deposit (1) 26 May 1994 (2) 26 May 1994	Accession Number (1) 69632 (2) 69633
<b>C. ADDITIONAL INDICATIONS</b> (leave blank if not applicable) <span style="float: right;">This information is continued on an additional sheet <input type="checkbox"/></span>	
Depositor: Molecular Biology Resources, Inc. These deposits were made pursuant to the provisions of the Budapest Treaty Host strains: (1) E coli DH5 $\alpha$ ' (2) E coli DH5 $\alpha$ F' IQ	
<b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> (if true indications are not for all designated States)	
<b>E. SEPARATE FURNISHING OF INDICATIONS</b> (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application C. Williams Authorized officer PCT International Division	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer



**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : C12N 15/00, 15/52, 15/70, C07K 14/195, C12N 1/21	A3	(11) International Publication Number: <b>WO 96/14405</b> (43) International Publication Date: 17 May 1996 (17.05.96)
(21) International Application Number: PCT/US95/15327 (22) International Filing Date: 3 November 1995 (03.11.95)	<b>(81) Designated States:</b> AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> <i>With an indication in relation to a deposited microorganism furnished under Rule 13bis separately from the description.</i> <i>Date of receipt by the International Bureau:</i> 29 December 1995 (29.12.95)	
(30) Priority Data: 08/334,645 4 November 1994 (04.11.94) US		
(71) Applicant: MOLECULAR BIOLOGY RESOURCES, INC. [US/US]; 5520 W. Burleigh Street, Milwaukee, WI 53210 (US).		
(72) Inventors: MUELLER, Reinhold, D.; 1930 North 86th Street, Wauwatosa, WI 53226 (US); SKOWRON, Piotr, M.; 3533 Salerno Court #8, Middleton, WI 53562 (US); SWAMI-NATHAN, Neela; 3026 Bosshard Drive, Madison, WI 53711 (US); PIEHL, Richard, F.; 2317 East Menlo Boulevard, Shorewood, WI 53211 (US).		
(74) Agents: POCHOPIEN, Donald, J. et al.; 6300 Sears Tower, 233 South Wacker Drive, Chicago, IL 60606-6402 (US).	(88) Date of publication of the international search report: 17 October 1996 (17.10.96)	

**(54) Title:** BIOLOGICALLY ACTIVE FRAGMENTS OF THERMUS FLAVUS DNA POLYMERASE**(57) Abstract**

The present invention is directed to a DNA encoding a biologically active fragment of a thermostable, full length DNA polymerase I enzyme of *Thermus flavus*. More particularly, the invention is directed to a DNA encoding an approximately 63,000 dalton DNA polymerase that lacks 274 amino acids from the N-terminus of the approximately 94,000 dalton *T. flavus* DNA polymerase I, and to the protein encoded thereby which has been designated the *T. flavus* DNA polymerase I exo fragment. The enzyme fragments are useful in DNA sequencing, Thermal Cycle Labeling, Polymerase Chain Reaction, and other molecular biological applications.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## INTERNATIONAL SEARCH REPORT

In national Application No.  
PCT/US 95/15327A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 C12N15/00 C12N15/52 C12N15/70 C07K14/195 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO,A,94 29482 (THIRD WAVE TECHNOLOGIES, INC.) 22 December 1994 see page 9, line 18 - page 10, line 22 see page 29, line 20 - page 32, line 2 see SEQ ID NO:6 see figure 2; example 2 ---	1,6-9, 12-14,16
X	WO,A,91 09950 (CETUS CORPORATION) 11 July 1991	1,6-9, 12-14,16
Y	see page 2, line 18 - page 3, line 37  see page 5, line 20 - page 18, line 30 see page 22, line 3 - page 23, line 10 see examples 1,2,7 ---	2-5,10, 11,15, 17-25
	--- -/--	

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents :

- \* "A" document defining the general state of the art which is not considered to be of particular relevance
- \* "E" earlier document but published on or after the international filing date
- \* "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \* "O" document referring to an oral disclosure, use, exhibition or other means
- \* "P" document published prior to the international filing date but later than the priority date claimed

\* "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\* "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\* "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\* "&" document member of the same patent family

Date of the actual completion of the international search

14 August 1996

Date of mailing of the international search report

03.09.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+ 31-70) 340-3016

Authorized officer

Donath, C

# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/US 95/15327

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,92 06200 (CETUS CORPORATION) 16 April 1992	1,6-9, 12-14,16
Y	see page 5, line 33 - page 10, line 27	2-5,10, 11,15, 17-25
	see page 18, line 4 - line 23 see page 28, line 18 - page 36, line 32 SEQ ID NO:9 and 10	
X	--- J.FERMENT.BIOENG., vol. 76, no. 4, 1993, pages 265-269, XP002010862 ASAKURA,K. ET AL.: "Cloning, nucleotide sequence, and expression in Escherichia coli of DNA polymerase gene (polA) from Thermus thermophilus HB8"	1,6-9, 12,16
Y	see the whole document	2-5,10, 11,13-15
Y	--- WO,A,92 03556 (CETUS CORPORATION) 5 March 1992 see page 2, line 28 - page 3, line 16 see page 10, line 1 - line 37 see page 14, line 10 - line 17 see examples 1-3,5	1-25
Y	--- US,A,5 352 778 (NEW ENGLAND BIOLABS, INC.) 4 October 1994 see column 2, line 60 - column 3, line 34 see column 5, line 6 - line 58 see column 9, line 31 - column 15, line 14 see examples 2,15,16	1-25
Y	--- BIOCHEMISTRY (A TRANSLATION OF BIOKHEMIYA), vol. 46, no. 9, September 1981, pages 1247-1254, XP002010863 KALEDIN,A.S. ET AL.: "Isolation and properties of DNA polymerase from the extremely thermophilic bacterium Thermus flavus" see the whole document	1-16
Y	--- NUCLEIC ACIDS RESEARCH, vol. 20, no. 21, 1992, page 5839 XP002010864 AKHMETZJANOV,A.A. AND VAKHITOV,V.A.: "Molecular cloning and nucleotide sequence of the DNA polymerase gene from Thermus flavus" see the whole document	1-16
	-----	

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.  
PCT/US 95/15327

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO-A-9429482	22-12-94	US-A-	5541311	30-07-96
		AU-B-	7052094	03-01-95
		CA-A-	2163015	22-12-94
		EP-A-	0711361	15-05-96
-----				
WO-A-9109950	11-07-91	AU-B-	6329694	01-09-94
		AU-B-	646430	24-02-94
		AU-B-	7176491	24-07-91
		AU-B-	656315	02-02-95
		AU-B-	7244491	24-07-91
		CA-A-	2071196	23-06-91
		CA-A-	2071213	23-06-91
		EP-A-	0506825	07-10-92
		EP-A-	0506889	07-10-92
		JP-T-	5504887	29-07-93
		JP-T-	5505105	05-08-93
		US-A-	5407800	18-04-95
		US-A-	5310652	10-05-94
		WO-A-	9109944	11-07-91
-----				
WO-A-9206200	16-04-92	AU-B-	4086896	26-04-96
		AU-B-	663474	12-10-95
		AU-B-	8668891	28-04-92
		CA-A-	2090614	29-03-92
		EP-A-	0550687	14-07-93
		US-A-	5405774	11-04-95
		US-A-	5466591	14-11-95
		US-A-	5455170	03-10-95
		AU-B-	658378	13-04-95
		AU-B-	8907791	28-04-92
		CA-A-	2092317	29-03-92
		EP-A-	0550696	14-07-93
		JP-T-	6504196	19-05-94
		WO-A-	9206202	16-04-92
-----				
WO-A-9203556	05-03-92	US-A-	5374553	20-12-94
		AU-B-	653747	13-10-94
		AU-B-	8501491	17-03-92
		CA-A-	2089495	14-02-92
		DE-T-	544789	16-12-93

# INTERNATIONAL SEARCH REPORT

Information on patent family members

L. International Application No  
PCT/US 95/15327

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9203556		EP-A-	0544789
		JP-A-	7147990
		JP-B-	7108220
		JP-T-	6500020
		US-A-	5420029
-----			
US-A-5352778	04-10-94	US-A-	5500363
		US-A-	5210036
		US-A-	5322785
		DE-T-	455430
		EP-A-	0455430
		JP-A-	5068547
		DE-T-	547920
		EP-A-	0547920
		JP-A-	6007160
-----			